

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:34:12 ; Search time 26.49 Seconds  
(without alignments)  
51.633 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209  
Sequence: 1 A1SYGNSADAPYVAKIGOVDAKQINKNATVGIYGYN 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_36:\*

1: /SIDSL/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
2: /SIDSL/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
3: /SIDSL/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
4: /SIDSL/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
5: /SIDSL/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
6: /SIDSL/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
7: /SIDSL/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
8: /SIDSL/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
9: /SIDSL/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
10: /SIDSL/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
11: /SIDSL/gcgdata/geneseq/geneseqp/AA1990.DAT:\*  
12: /SIDSL/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDSL/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDSL/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDSL/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDSL/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDSL/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDSL/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDSL/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDSL/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDSL/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	209	100.0	40	21	Y84611 N-terminal peptide
2	209	100.0	180	21	Y84612 The outer membrane
3	152	72.7	31	21	Y56863 B. catarrhalis 20
4	57	27.3	680	21	Y79196 Maize glutamine:fr
5	53	25.4	1235	17	W21572 Rat brain serine e
6	52	24.9	1249	17	W21571 Rat brain homology
7	52	24.9	199	14	R34398 Helicobacter pylori
8	51	24.4	1074	20	Y05734 Human Grb7 effecto
9	50.5	24.2	221	20	Y37398 Protein involved i
10	50	23.9	833	18	W27148 HMG-CoA reductase
11	49.5	23.7	255	18	W55477 H. pylori ORF 07ep
12	49.5	23.7	255	18	W55201 H. pylori ORF 01cp

13	49.5	23.7	255	20	Y17205	H. pylori outer me
14	49.5	23.7	1203	21	Y83275	Candida albicans C
15	49.5	23.7	1203	21	Y84814	Amino acid sequenc
16	49	23.4	540	15	R48667	Chitinase 1, Rhiz
17	48.5	23.2	435	14	R32999	Rat choline kinase
18	48.5	23.2	652	21	Y83039	Cry9a toxin N-ter
19	48.5	23.2	656	19	Y83040	Synthesized Cry9a
20	48.5	23.2	1156	20	Y46856	Bacillus thuringie
21	48.5	23.2	1156	20	Y24959	Bacillus thuringie
22	48.5	23.2	1279	21	Y82988	Native Cry9a toxin
23	48	23.0	341	16	R72964	Pig kidney cell mu
24	48	23.0	341	16	R70142	Porcine mutarotase
25	48	23.0	532	15	R47237	Wild-type feline H
26	48	23.0	532	17	R96588	Feline herpesvirus
27	48	23.0	532	19	W80432	Feline herpesvirus
28	48	23.0	532	20	Y32475	DNA encoding feli
29	47.5	22.7	283	20	Y33687	S. erythraea PRU37
30	47.5	22.7	629	21	Y57585	Human papillomavir
31	47.5	22.7	2514	21	Y75097	Neisseria meningit
32	47	22.5	127	20	Y26084	Helicobacter pylor
33	47	22.5	541	19	W44165	Entodinium candida
34	46.5	22.2	507	6	P50263	Sequence encoded b
35	46	22.0	165	19	W38713	S. pneumoniae glut
36	46	22.0	493	15	R48659	Chitinase derivati
37	46	22.0	778	19	W35004	Thermotoga maritim
38	46	22.0	832	20	W74089	Human HPT-1-protei
39	46	22.0	1287	16	R79944	Helicobacter pylor
40	46	22.0	1291	16	R75201	Tyrosine phosphata
41	46	22.0	1296	14	R41198	CT. Helicobacter
42	46	22.0	2893	19	W98828	H. pylori GPO 148
43	46	22.0	2893	19	W71556	Helicobacter poly

44	45.5	21.8	218	21	Y44507	C-terminal fragmen
45	45.5	21.8	251	16	R80192	Minor ampullate sp
ALIGNMENTS						
RESULT 1						
ID	Y84611	standard; peptide: 40 AA.				
AC	Y84611;					
DE	25-JUL-2000	(first entry)				
DF						
DE		N-terminal peptide of outer membrane protein 21 of strain ATCC49143.				
DE		Outer membrane protein 21: OMP21; strain ATCC49143; Bacterial infection;				
KW		otitis media; respiratory infection; sinusitis; pneumonia; immunisation.				
XX						
OS		Moraxella catarrhalis.				
XX						
PN		WO200018910-A1.				
XX						
PD		06-APR-2000.				
XX						
PF		01-OCT-1999; 99WO-US22918.				
XX						
PR		01-OCT-1998; 98US-0164714.				
XX						
PA		(ANTE-) ANTEX BIOLOGICS INC.				
PI		Tucker K, Tillmann UF;				
XX						
DR		WPI; 2000-293149/25.				
XX						
PT		Isolated outer membrane protein from a Moraxella catarrhalis strain				
PT		used for diagnosis treatment and prevention of disease caused by M.				
XX		catarrhalis e.g. pneumonia, otitis media and respiratory infections -				
PS		Claim 2; Page 63; 108pp; English.				
XX						

CC The present sequence represents the N-terminal peptide of a mature  
CC outer membrane protein 21 (OMP21) of Moraxella catarrhalis strain  
CC ATCC49143. The specification describes an OMP21 from a Moraxella  
CC catarrhalis strain, which has an apparent molecular weight of 16-20 kD  
CC as determined by sodium dodecylsulfate polyacrylamide gel electrophoresis  
CC (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in  
CC prophylactic and therapeutic compositions for treating a M. catarrhalis  
CC bacterial infection, otitis media, respiratory infections, sinusitis  
CC and pneumonia. They are useful as reagents for the clinical or medical  
CC diagnosis of M. catarrhalis infections and for scientific research on  
CC the properties of pathogenicity, virulence and infectivity of  
CC M. catarrhalis and host defence mechanisms. The antibodies, particularly  
CC those that are cytotoxic may be used in passive immunisation to prevent  
CC or attenuate M. catarrhalis infections of animals e.g. humans.

XX  
SQ Sequence 40 AA;

Query Match 100.0%; Score 209; DB 21; Length 40;  
Best Local Similarity 100.0%; Pred. No. 8.9e-24;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNTAYGTYAGYN 40  
Db 1 aisynsadaqpyvgakigvdakqngkntaygiyagyn 40  
|||||

RESULT 2  
Y84612 Y84612 standard; Protein; 180 AA.  
XX  
AC Y84612;  
XX  
DT 25-JUL-2000 (first entry)  
XX  
DE The outer membrane protein 21 of strain ATCC49143.  
XX  
KW Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;  
KM otitis media; respiratory infection; sinusitis; pneumonia; immunisation.  
XX  
OS Moraxella catarrhalis.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 71 /note="Ala encoded by CCC"  
XX  
XX  
PN WO200018910-A1.  
XX  
PD 06-APR-2000.  
XX  
PF 01-OCT-1999; 99WO-US22918.  
XX  
PR 01-OCT-1998; 98US-0164714.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Tucker K, Tillmann UF;  
XX  
XX  
DR WPI: 2000-293149/25.  
DR N-PSDB; A12591.  
XX  
XX  
PT Isolated outer membrane protein from a Moraxella catarrhalis strain  
PT used for diagnosis treatment and prevention of disease caused by M.  
PT catarrhalis e.g. pneumonia, otitis media and respiratory infections -  
XX  
PS  
XX Claim 2; Fig 4; 108pp; English.

CC The present sequence represents an outer membrane protein 21 (OMP21)  
CC of Moraxella catarrhalis strain ATCC49143. The OMP21 protein has an  
CC apparent molecular weight of 16-20 kD as determined by sodium  
CC dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21,  
CC its nucleic acids and antibodies can be used in prophylactic and  
CC therapeutic compositions for treating a M. catarrhalis bacterial

CC infection, otitis media, respiratory infections, sinusitis and  
CC pneumonia. They are useful as reagents for the clinical or medical  
CC diagnosis of M. catarrhalis infections and for scientific research on  
CC the properties of pathogenicity, virulence and infectivity of  
CC M. catarrhalis and host defence mechanisms. The antibodies, particularly  
CC those that are cytotoxic may be used in passive immunisation to prevent  
CC or attenuate M. catarrhalis infections of animals e.g. humans.

XX  
SQ Sequence 180 AA;

Query Match 100.0%; Score 209; DB 21; Length 180;  
Best Local Similarity 100.0%; Pred. No. 5.6e-23;  
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNTAYGTYAGYN 40  
Db 23 aisynsadaqpyvgakigvdakqngkntaygiyagyn 62  
|||||

RESULT 3  
Y56863 Y56863 standard; peptide; 31 AA.  
XX  
AC Y56863;  
XX  
DT 10-APR-2000 (first entry)  
XX  
DE B. catarrhalis 20 kD antigen N-terminal fragment.  
XX  
KW Branhamella catarrhalis; Moraxella catarrhalis; antigen; vaccine;  
KM immunogenic; respiratory tract infection; otitis media.  
XX  
OS Branhamella catarrhalis.  
XX  
PN WO9958563-A2.  
XX  
PD 18-NOV-1999.  
XX  
PF 11-MAY-1999; 99WO-GB01473.  
XX  
PR 11-MAY-1998; 98GB-0010084.  
XX  
PA (CORT-) CORTECS UK LTD.  
XX  
PI Cripps AW, Kyd J;  
XX  
XX  
DR WPI: 2000-116286/10.  
XX  
XX  
PT Novel antigens of Branhamella catarrhalis used for diagnosis, detection  
PT and in vaccines -  
XX  
PS  
XX Claim 5; Page 22; 32pp; English.

CC The invention relates to novel Branhamella catarrhalis (Moraxella  
CC catarrhalis) antigens having an apparent molecular weight of about 14-71  
CC kDa (as determined by SDS- PAGE). The antigens can be used to prepare  
CC vaccines and immunogenic compositions for the treatment and prophylaxis  
CC of Branhamella catarrhalis infections, respiratory tract infections, and  
CC otitis media. Antibodies against the antigens can be used for diagnosis  
CC and purification of the antigens. Sequences Y56863-871 represent peptide  
CC peptide fragments from the B. catarrhalis antigens of the invention.

XX  
SQ Sequence 31 AA;

Query Match 72.7%; Score 152; DB 21; Length 31;  
Best Local Similarity 96.8%; Pred. No. 1.4e-15;  
Matches 30; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AISYNSADAQPYVGAKIGVDAKQINGKNT 31  
Db 1 aisynsadaqpyvgakigvdakqlnknt 31  
|||||

Sequence	Accession	Score	DB	Length	680
Query Match	27.38	Score 57	DB 21	Length 680	
Best Local Similarity	32.08	Pred. No. 4.4			
Matches 16	Conservative % 9	Mismatches 7	Indels 18	Gaps 103	
6 NSADADPVVGA-----K-----GVDAKQJNGKNTAYGCIYAG 38					
55 sssdarpyagppjlvfgeqkfenlvrsyvsevdakvn-ldaaafsvhag 103					

DT	20-AUG-1997 (first entry)
XX	
DE	Rat brain serine ectopeptidase.
XX	
KW	Serine ectopeptidase; tripeptidylpeptidase II; "pp II;
KM	alternative splicing; rodent; rat; cerebral cortex;
KW	neuropeptide; cholecystokinin; CCK; inactivation; degradation;
KW	anorexia; schizophrenia; Parkinson's disease; depression;
KW	irritable bowel syndrome; bulimia; pathological obesity.
XX	
OS	Rattus sp.
XX	
FT	Key Location/Qualifiers
FT	Region 85..1235
FT	/note= "The C-terminal region of the protein encoded by the alternatively spliced clone has a sequence identical to this region"
XX	
PN	W09635805-A2.
XX	
PD	14-NOV-1996.
XX	
PF	09-MAY-1996; 96WO-FR00700.
XX	
PR	09-MAY-1995; 95FR-0005489.
XX	
PA	(INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX	
PI	Bambal RB, Bishop PB, Bourgeat P, Chan S, Ganellin CR;
PI	Léhlond B, Moore AMJ, Schwartz JC, Vargas F;
PI	Lihua Z, Rose C;
XX	
DR	WP1: 1996-518693/51.
DR	N-PSDB: T72654.
XX	
PT	Screening medicaments for treating disorders linked to inactivation of endogenous neuro-peptide(s) - by contacting candidate molecule with membrane trl:peptidyl-peptidase (homologue) and measuring enzyme activity
XX	
PS	Claim 4: Pages 166-171; 212pp; French.
XX	
CC	An enzyme with specificity for cholecystokinin (CCK) substrates (specifically the non-sulphated CCK8 and the CCK5 peptides) was purified from rat cerebral cortex membranes using high performance liquid chromatography. Based on amino acid sequence data from the purified enzyme, PCR primers were designed to amplify oligonucleotide probes A and B of lengths 350 and 380 nucleotides, respectively. Using these probes, two distinct clones were identified in a rat brain cDNA library. The first clone encoded the protein which is the rodent homologue of human tripeptidylpeptidase II. The second clone differed in the 5' region, probably as a result of alternative splicing, and encodes a protein of the present sequence which appears to be a serine ectopeptidase. The rat enzymes are preferred for use in a new method of screening for medicaments for treating disorders linked to the inactivation or degradation (or being treatable by reversing physiological degradation) of endogenous neuropeptides. In the method, a candidate molecule is contacted with a membrane tripeptidylpeptidase or homologue, and enzyme activity is measured. Disorders linked to inactivation or degradation of endogenous neuropeptides, include food intake disorders, cognitive and motor disorders such as anorexia, schizophrenia, Parkinson's disease and depression, as well as gastrointestinal transit disorders such as irritable bowel syndrome, bulimia and pathological obesity.
XX	
SO	Sequence 1235 AA;

Query Match                      25.4%: Score 53; DB 17; Length 1235;  
Best Local Similarity    36.7%; Pred No. 35;  
Matches    11; Conservative    6; Mismatches    13; Indels       0; Gaps       0;

10 ADPYGAKIGVDARQINCKNTATYGIACY 39

Db 1124 aqphdgaagdaakeegestleslsety 1153

```

||||| 1 1 :||| 1:1 1
RESULT 6
W21571 W21571 standard; Protein; 1249 AA.
XX
AC W21571:
XX
DT 20-AUG-1997 (first entry)
XX
DE Rat brain homologue of human tripeptidylpeptidase II.
XX
KW Tripeptidylpeptidase II; TRP II; rodent; rat; cerebral cortex;
KW neuropeptide; cholecystokinin; CCK; inactivation; degradation;
KW anorexia; schizophrenia; Parkinson's disease; depression;
KW irritable bowel syndrome; bulimia; pathological obesity;
KW alternative splicing.
XX
OS Rattus sp.
XX
FH Key Location/Qualifiers
FT Region 99..1249
FT /note- "The C-terminal region of the protein encoded
FT by the alternatively spliced clone has a
FT sequence identical to this region"
XX
PN M09635805-A2.
XX
PD 14-NOV-1996.
XX
PE 09-MAY-1996; 96WO-FR00700.
XX
PR 09-MAY-1995; 95FR-0005489.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bambaï RB, Bishop PB, Bourgeat P, Chan S, Gannellini CR;
PI Loblond B, Moore ANJ, Schwartz JC, Vargas F;
PI Lihua Z, Rose C;
XX
DR WPI: 1996-518693/51.
DR N-PSDB: T72653.
XX
PT Screening medicaments for treating disorders linked to inactivation
PT of endogenous neuro-peptide(s) by contacting candidate molecule
PT with membrane tri-peptidyl-peptidase (homologue) and measuring
PT enzyme activity
XX
PS Claim 3; Pages 161-165; 212pp: French.
XX
CC An enzyme with specificity for cholecystokinin (CCK) substrates
CC (specifically the non-sulphated CCK8 and the CCK5 peptides) was
CC purified from rat cerebral cortex membranes using high performance
CC liquid chromatography. Based on amino acid sequence data from the
CC purified enzyme, PCR primers were designed to amplify oligonucleotide
CC probes A and B of lengths 350 and 380 nucleotides, respectively.
CC Using these probes, two distinct clones were identified in a rat
CC brain cDNA library. The first clone encoded the protein of the
CC present sequence which is the rodent homologue of human tripeptidyl-
CC peptidase II. The second clone differed in the 5' region, probably
CC as a result of alternative splicing, and appears to be a serine
CC ectopeptidase. The rat enzymes are preferred for use in a new
CC method of screening for medicaments for treating disorders linked
CC to the inactivation or degradation (or being treatable by retarding
CC physiological degradation) of endogenous neuropeptides. In the
CC method, a candidate molecule is contacted with a membrane tripeptidyl-
CC peptidase or homologue, and enzyme activity is measured.
CC Disorders linked to inactivation or degradation of endogenous
CC neuropeptides, include food intake disorders, cognitive and motor
CC disorders such as anorexia, schizophrenia, Parkinson's disease and
CC depression, as well as gastrointestinal transit disorders such as

```

CC irritable bowel syndrome, bulimia and pathological obesity.  
 XX  
 SQ Sequence 1249 AA;

Query Match 25.4%; Score 53; DB 17; Length 1249;  
 Best Local Similarity 36.7%; Pred. No. 36;  
 Matches 11; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 10 AOPYVGAKIGQVDAKQINGKNPAGYVAGY 39  
 ||| 1 1 :||| 1:1 1  
 Db 1138 aqphdgaagdaakeegestleslsety 1167

```

RESULT 7
R34398 R34398 standard; Protein; 199 AA.
XX
AC R34398:
XX
DT 03-AUG-1993 (first entry)
XX
DE Helicobacter pylori ureg gene.
XX
KW ureG; uref; ureH; ureI; stomach; gastric mucosa; gastric ulcer;
KW duodenal ulcer; acid tolerance; urease operon.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT Region 8..15
FT /note- "ATP-binding"
XX
PN M09307273-A.
XX
PD 15-APR-1993.
XX
PE 02-OCT-1992; 92WO-FR00921.
XX
PR 03-OCT-1991; 91FR-0012198.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MED.
XX
PI (INSP ) INST PASTEUR.
XX
PI Cussac V, Ferrero R, Labigne A;
XX
DR WPI: 1993-134462/16.
DR N-PSDB: Q40312.
XX
PT Helicobacter pylori genes useful in diagnosis, vaccines and
PT treatment - necessary for the regulation and maturation of urease
PT
XX
PS Claim 18; Fig 4; 94pp: French.
XX
CC The ureg gene is one of 5 new urease genes to be identified by
CC deletion studies in E. coli where all 5 were found to be necessary
CC for functional urease expression (regulation/maturation). The ureg
CC gene encodes a protein of estimated mol. wt. 21.7kD. The ureg
CC polypeptide of H. pylori has 92% conservation and 59% identity with
CC the ureg polypeptide of K. aerogenes. In K. aerogenes the ureg gene
CC is implicated in the activation of the apoenzyme by incorporation
CC of nickel in the urease subunits. See also R34395-R34399.
XX
SQ Sequence 199 AA;

```

Query Match 24.9%; Score 52; DB 14; Length 199;  
 Best Local Similarity 40.0%; Pred. No. 5.3;  
 Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;  
 QY 6 NSADAPYVGA--KIGQVDAKQINGKNYAY 33  
 ||| 1 1 :||| 1:1 1  
 Db 145 nkldlapyvgadlkvrmdskkkaaksp1f 174

```

RESULT 8
ID Y05734 standard; Protein: 1074 AA.
XX
AC Y05734;
XX
DE 19-JUL-1999 (first entry)
XX
DE Human Grb7 effector 2.2412 protein.
XX
KW Grb7 effector; 2.2412 protein; human; signal transduction;
KW tumour marker; breast cancer; prostate cancer; prognosis;
KW diagnosis.
XX
OS Homo sapiens.
XX
PN W09915647-A1.
XX
PD 01-APR-1999.
XX
PE 23-SEP-1998; 98WO-AU00795.
XX
PR 23-SEP-1997; 97AU-0009388.
XX
PA (GARV-) GARVAN INST MEDICAL RES.
XX
PI Daly RJ, Sutherland RL;
XX
DR WPI: 1999-254707/21.
DR N-PSDB; X25366.
XX
PT New candidate effector for the Grb7 family of signaling proteins,
XX and specific antibody, useful for detection and treatment of cancer
XX
XX Claim 9; Fig 1; 24pp; English.
XX
XX The present sequence represents a novel candidate effector for
XX the Grb7 family of signaling proteins, termed 2.2412. The
XX sequence is predicted from a partial cDNA (see X25366). 2.2412
XX cDNA transcripts (7 kb) were detected in all tissues examined
XX with the exception of kidney. Expression was particularly high in
XX skeletal muscle and placenta. Analysis of the sequence revealed
XX significant homology to a large number of proteins containing
XX ankyrin-like repeats. The 2.2412 gene was localised to between
XX chromosome 10q23.2 and proximal 10q23.32. Deletions in the
XX 10q22-25 region have been detected in human breast, prostate,
XX renal, small cell lung and endometrial carcinomas, glioblastoma
XX multiforme, melanoma and meningiomas. Detection of the protein
XX encoded by the 2.2412 cDNA in a sample should provide a useful
XX tumour marker and/or prognostic indicator for certain human
XX cancers, in particular breast cancer and prostate cancer.
XX Antagonism of the interaction between Grb7 family members and the
XX encoded protein should provide a novel treatment strategy for human
XX diseases exhibiting aberrant receptor tyrosine kinase signalling,
XX such as cancer. Anti-Grb7 antibodies can be used in methods of
XX detecting the presence of 2.2412 protein in a sample.
XX
SQ Sequence 1074 AA:

```

```

Query Match          24.4%; Score 51; DB 20; Length 1074;
Best Local Similarity 35.8%; Pred. No. 58;
Matches 19; Conservative 4; Mismatches 16; Indels 14; Gaps 3;
QY 2 ISYGS-ADAGPYVGAKIGVD-----AKQINGK-NTAVGIAGYN 40
   |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 393 ISYGSADAGPYVGAKIGVD-----AKQINGK-NTAVGIAGYN 445

```

```

ID Y37398 standard; Protein: 221 AA.
XX
AC Y37398;
XX
DE 07-OCT-1999 (first entry)
XX
DE Protein involved in intermediate metabolism of sugars and/or cofactors.
XX
KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital diseases; peritrichiasis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW bartolinitis; pneumopathy; venereal lymphogranulomatosis.
XX
OS Chlamydia trachomatis.
XX
PN W09928475-A2.
XX
PD 10-JUN-1999.
XX
PE 27-NOV-1998; 98WO-IB01939.
XX
PR 04-NOV-1998; 98US-0107077.
XX
PR 28-NOV-1997; 97FR-0015041.
XX
PR 17-DEC-1997; 97FR-0016034.
XX
PA (GEST ) GENSET.
XX
PI Griffois R;
XX
DR WPI: 1999-371125/31.
XX
PT Genome sequence of Chlamydia trachomatis
XX
XX Disclosure: Page 1105-1106; 1755pp; English.
XX
XX Y36754-Y37949 are encoded by open reading frames (ORFs) of the genome of
XX Chlamydia trachomatis (see Z01425). The polypeptides can be used as
XX vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
XX can also be used to control growth of the microorganism. Chlamydia
XX trachomatis is responsible for a large number of diseases, e.g. eye
XX diseases such as conventional trachoma, nonendemic trachoma,
XX paratrachoma, and inclusion conjunctivitis; genital diseases such as
XX nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
XX peritrichiasis, bartolinitis; pneumopathy in breast feeding infants;
XX and venereal lymphogranulomatosis. The polypeptides of the invention
XX may be of use in treating these diseases.
XX
SQ Sequence 221 AA:

```

```

Query Match          24.2%; Score 50.5; DB 20; Length 221;
Best Local Similarity 38.2%; Pred. No. 9.9;
Matches 13; Conservative 3; Mismatches 15; Indels 3; Gaps 1;
QY 5 GNSADQPYVG---AKIGVDKQINGKNTAVGI 35
   |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 99 GNSADQPYVG---AKIGVDKQINGKNTAVGI 132

```

```

RESULT 10
ID W27148 standard; Protein: 833 AA.
XX
AC W27148;
XX
DE 04-DEC-1997 (first entry)
XX
DE HMG-CoA reductase degradation protein Hrd3p.
XX
DE 3-hydroxy-3-methylglutaryl: coenzyme A; cholesterol; Hrd1p; Hrd2p;
KW Hrd3p; hypercholesterolaemia; yeast.
XX
OS Saccharomyces cerevisiae.

```

[illegible]

XX	27-MAR-1997;	97WO-0505223.	
PF			
XX	06-DEC-1996;	96US-0761318.	
PR	29-MAR-1996;	96US-0625811.	
PR	02-APR-1996;	96US-0758731.	
PR	25-OCT-1996;	96US-0736905.	
XX	28-OCT-1996;	96US-0738859.	
PA	(ASPR ) ASPRA AB.		
XX			
PI	Alm RA, Smith D;		
XX			
DR	WPI, 1997-503122/46.		
XX	N-PSDB; V24886.		
XX			
PT	Helicobacter pylori nucleic acid sequences and encoded		
PT	polypeptide(s) - useful in vaccines to treat or prevent H. pylori		
XX	infection and for diagnosis of H. pylori infection		
PS			
XX	Claims 14,80; Pages 684-685; 1145pp; English.		
CC			
CC	This sequence is a H. pylori cell envelope outer membrane		
CC	protein (OMP) having a C-terminal tyrosine cluster motif.		
CC	The protein may be used in a vaccine to prevent or treat H. pylori		
CC	infection or to identify H. pylori polypeptide binding compounds,		
CC	useful as potential H. pylori life cycle activators or inhibitors.		
CC	The DNA and probes derived from it may be used for the		
CC	identification of H. pylori in a sample, and the diagnosis of		
CC	H. pylori infection. Nucleic acid sequences complementary to the		
CC	DNA act as antisense sequences, and can be used to prevent the		
CC	translation of H. pylori mRNA. Antibodies against the protein can		
CC	be used in immunoassays to evaluate the abundance and distribution		
CC	of H. pylori-specific antigens. The genomic sequence of H. pylori		
CC	(ATCC 55679) was determined from overlapping contigs generated		
CC	by mechanically shearing the bacterial DNA. The sequences were		
CC	analysed for ORF of at least 180 nucleotides, and the predicted		
CC	coding regions defined by computer evaluation. To identify likely		
CC	H. pylori antigens for vaccine development, the amino acid		
CC	sequences predicted from various ORF were analysed for significant		
CC	homology to other known or exported membrane proteins. Having		
CC	identified and determined the sequences of interest, particular		
CC	regions can be isolated from H. pylori by PCR amplification for		
CC	recombinant polypeptide production, e.g. in E. coli hosts.		
XX			
XX	Sequence 255 AA:		
SO			
	Query Match 23.7%; Score 49.5; DB 18; Length 255;		
	Best Local Similarity 34.6%; Pred. No. 16;		
	Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5		
QY	3 SYG-----NSADQPYVGAKTIGVD-AKQIN-----GKMTAVGIV---AGYN 40		
	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :		
DB	101 tgygysynhans-ftvsgklgimdasqynmftgygdalnfyfeskyn 151		
RESULT 12			
W55201	W55201 standard; Protein: 255 AA.		
XX			
AC	W55201;		
XX			
XX	15-JUN-1998 (first entry)		
DE			
XX	H. pylori ORF Olcpl14Iorf2 protein.		
XX			
KM	Cytoplasmic; vaccine; prevention; treatment; infection; envelope;		
KM	identification; binding compound; bacteria; life cycle; activator;		
XX	inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.		
OS	Helicobacter pylori.		
XX			

```

PN      MW0737044-A1.
XX
XX      09-OCT-1997.
XX
XX      27-MAR-1997; 97WO-US05223.
XX
XX      06-DEC-1996; 96US-0761318.
XX      29-MAR-1996; 96US-0625811.
XX      02-APR-1996; 96US-0758731.
XX      25-OCT-1996; 96US-0736905.
XX      28-OCT-1996; 96US-0738859.
XX
XX      (ASTR ) ASTRA AB.
PA
PA      Alm RA, Smith D;
PI
XX      WPI: 1997-503122/46.
XX
XX      N-PSDB: V24610.
DR
XX
XX      Helicobacter pylori nucleic acid sequences and encoded
XX      PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
XX      PT infection and for diagnosis of H. pylori infection
XX
XX      Claim 14; Page 453; 1145pp; English.
PS
XX
XX      This sequence is a H. pylori protein of unspecified function.
XX      CC The protein may be used in a vaccine to prevent or treat H. pylori
XX      CC infection or to identify H. pylori polypeptide binding compounds,
XX      CC useful as potential H. pylori life cycle activators or inhibitors. The
XX      CC DNA and probes derived from it may be used for the identification of
XX      CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
XX      CC acid sequences complementary to the DNA act as antisense sequences and
XX      CC can be used to prevent the translation of H. pylori mRNA. Antibodies
XX      CC against the protein can be used in immunoassays to evaluate the abundance
XX      CC and distribution of H. pylori-specific antigens. The genomic sequence of
XX      CC H. pylori (ATCC 55679) was determined from overlapping contigs generated
XX      CC by mechanically shearing the bacterial DNA. The sequences were analysed
XX      CC for ORF of at least 180 nucleotides, and the predicted coding regions
XX      CC defined by computer evaluation. To identify likely H. pylori antigens for
XX      CC vaccine development, the amino acid sequences predicted from various ORF
XX      CC were analysed for significant homology to other known or exported
XX      CC membrane proteins. Having identified and determined the sequences of
XX      CC interest, particular regions can be isolated from H. pylori by PCR
XX      CC amplification for recombinant polypeptide production, e.g. in E. coli
XX      CC hosts.
XX
XX      Sequence 255 AA;
SO
XX
XX      Query Match 23.7%; Score 49.5; DB 18; Length 255;
XX      Best Local Similarity 34.6%; Pred. No. 16;
XX      Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5
XX
Oy      3 SYG-----NSADAQPYVGAKIGVD-AKQIN-----GKNTAYGIV---AGYN 40
XX      :11 11:11:11:11:11:11:11:11:11:11:11:11:11:11:11:11:11:11
XX      Db 101 tygyysynhanis-fvysklgimdgasynmttygvfdalnfyfeskyn 151
XX
XX      RESULT 13
XX      ID Y17205 Y17205 standard; Protein; 255 AA.
XX      AC Y17205;
XX      DT 03-AUG-1999 (first entry)
XX      DE H. pylori outer membrane polypeptide.
XX      KW Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
XX      cellular immune response.
XX      OS Helicobacter pylori.
XX

```

PN	M09921959-A2.
XX	
PD	06-MAY-1999.
XX	
PE	28-OCT-1998; 98MO-US22883.
XX	
PR	17-DEC-1997; 97US-0993001.
PR	28-OCT-1997; 97US-0959131.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
XX	
DR	WPI; 1999-326698/27.
DR	N-PSDB; X75824.
XX	
PT	Cellular vaccine against Helicobacter pylori
XX	
PS	Claim 7; Page 287-288; 352pp; English.
XX	
CC	The invention relates to a vaccine for preventing or treating infections by Helicobacter pylori. The vaccine contains at least one isolated H. pylori polypeptide, or its fragments, in a carrier, where the carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a nucleic acid encoding the H. pylori polypeptide. The vaccines induce humoral and cellular immune responses. The vaccines are used to treat or prevent infections by H. pylori. Sequences X75779 to X75837 represent nucleic acid sequences encoding H. pylori outer membrane polypeptides (OMPs) Y11260 to Y11218.
CC	
CC	
CC	
SO	Sequence 255 AA:
Query Match	23.7%; Score 49.5; DB 20; Length 255;
Best Local Similarity	34.6%; Pred. No. 16;
Matches 18; Conservative 8; Mismatches 11; Indels 15; Gaps 5	
OY	3 SYG-----NSADAQPYVGAKTIGVD-AKQIN-----GKNATAGYGV--AGYN 40 ::    :  :    :                  Db      101 Tygyssynhnans-fvysklgimdgasgvnftlygvgdalynfyeskegyn 151
RESULT 14	
ID	Y83275 standard; Protein; 1203 AA.
AC	Y83275;
XX	
DT	16-AUG-2000 (first entry)
XX	
DE	Candida albicans CSA1 surface protein.
XX	
KM	Yeast; infection; detection; treatment; vaccine; antigen; CSA1;
KW	probe; primer; vaginitis; immunoglobulin G; IgG.
XX	
OS	Candida albicans.
XX	
PM	CA2274984-A1.
XX	
PD	10-JAN-2000.
XX	
PF	09-JUL-1999; 99CA-2274984.
XX	
PR	10-JUL-1998; 98CA-2237134.
XX	
PA	(UYLA-) UNIV LAVAL.
XX	
PI	Bourbonnais Y, Deslauriers N;
XX	
DR	WPI; 2000-318497/28.
DR	N-PSDB; Z93782.
XX	
PT	Novel nucleic acid sequence encoding a Candida albicans surface antigen

PT useful for detection, prophylaxis and treatment of candidiasis  
XX Claim 5; Page 32-36; 50pp; English.  
CC The CSA1 surface antigen is unique to *Candida albicans* and so can  
CC be exploited for the treatment of *Candida albicans* infections.  
CC Nucleic acids, probes and primers specific to the CSA1 coding  
CC sequence can be used to detect *C. albicans* infections in patients  
CC by binding to *C. albicans* CSA1 DNA or mRNA in a patient sample. The  
CC CSA1 protein, vaccines comprising it or antibodies directed against  
CC it can be used to treat *C. albicans* infections. The virulence of  
CC infection by *C. albicans* is reduced by binding of the antibody to  
CC the surface antigen and masking it. Recurrent vaginitis in  
CC immunocompetent patients can be created by repeated administration  
CC of CSA1 antigen to stimulate the production of immunoglobulin G  
CC antibodies. The antibodies raised against CSA1 are specific for *C. albicans* CSA1 antigen but not to other yeast or mycelial antigens,  
CC making detection and treatment easier.  
XX  
SQ Sequence 1203 AA;

Query Match 23.7%; Score 49.5; DB 21: Length 1203;  
Best Local Similarity 48.3%; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

OY 3 SYGNSADAOPYVGAKIGVDKQINGKNT 31  
||||| ||| | | | | |  
Db 951 sygnstiaqpstsk---sdaasttgpit 976

RESULT 15

Y84814  
Y84814 standard; protein; 1203 AA.

AC Y84814;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the CSA1 gene of *Candida albicans*.

XX CSA1 gene; surface antigen; yeast; passive immunization;

KW *Candida albicans* infection; vaginal infection; systemic infection;

KW anti-fungal.

XX *Candida albicans*.

OS  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..17  
FT /note= "signal peptide"

FT Domain 42..143  
FT /note= "C-rich domain"

FT Domain 233..334  
FT /note= "C-rich domain"

FT Domain 403..504  
FT /note= "C-rich domain"

FT Domain 565..666  
FT /note= "C-rich domain"

FT Domain 750..851  
FT /note= "C-rich domain"

FT Modified-site 883  
FT /note= "putative N-glycosylation site"

FT Modified-site 893  
FT /note= "putative N-glycosylation site"

FT Modified-site 903  
FT /note= "putative N-glycosylation site"

FT Modified-site 913  
FT /note= "putative N-glycosylation site"

FT Modified-site 914  
FT /note= "putative N-glycosylation site"

FT Modified-site 927  
FT /note= "putative N-glycosylation site"

FT Modified-site 928  
FT /note= "putative N-glycosylation site"

FT Modified-site /note= "putative N-glycosylation site"  
FT 937  
FT /note= "putative N-glycosylation site"  
FT Modified-site 938  
FT /note= "putative N-glycosylation site"  
FT Modified-site 954  
FT /note= "putative N-glycosylation site"  
FT Peptide 1184..1203  
FT /note= "predicted GPI-anchoring determinant"

CA2237134-A1.

PD 10-JAN-2000.

XX 10-JUL-1998;

PF 10-JUL-1998; 98CA-2237134.

PR 10-JUL-1998; 98CA-2237134.

XX (UYLA-) UNIV LAVAL.

PI Deslauriers N, Bourbonnais Y;

DR WPI; 2000-293503/26.

XX New CSA1 gene coding for a *Candida albicans* surface antigen, useful for

PT diagnosing the yeast form of *Candida albicans* within biological samples

PT by immunoassay or polymerase chain reaction -

PS Disclosure; Fig 4; 36pp; English.

XX The present sequence represents a protein encoded by the CSA1 gene.

CC The protein is a *Candida albicans* surface antigen. The CSA1 gene is

CC an intronless single copy gene in *Candida*, and RNA transcripts can

CC be detected in exponentially growing yeast cells. CSA1 gene expression

CC is strongly increased upon induction of the mycelial growth phase.

CC The CSA1 gene, its protein and antibodies against it are useful for

CC diagnosing the yeast form of *Candida albicans* within biological samples

CC by immunoassay or polymerase chain reaction (PCR). The antibody is

CC useful for passive immunization against *Candida albicans* infection.

CC The antibody is also useful in the treatment of vaginal or systemic

CC infection by *Candida*. The protein is useful as a therapeutic target

CC in the development of anti-fungal agents.

XX  
SQ Sequence 1203 AA;

Query Match 23.7%; Score 49.5; DB 21: Length 1203;  
Best Local Similarity 48.3%; Pred. No. 1.1e+02;  
Matches 14; Conservative 0; Mismatches 12; Indels 3; Gaps 1;

OY 3 SYGNSADAOPYVGAKIGVDKQINGKNT 31  
||||| ||| | | | | |  
Db 951 sygnstiaqpstsk---sdaasttgpit 976

Search completed: March 6, 2001, 14:36:12  
Job time: 120 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:34:37 ; Search time 19.25 Seconds  
(without alignments)  
37.313 Million cell updates/sec

Title: US-09-164-714-1  
Perfect score: 209  
Sequence: 1 A1SYGNSADAPYVGAIGGVDAKQJNGKNVAYGIYAGYN 40

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2.6/ptodata/2/1aa/5A.COMB.pep:\*
- 2: /cgn2.6/ptodata/2/1aa/5B.COMB.pep:\*
- 3: /cgn2.6/ptodata/2/1aa/6.COMB.pep:\*
- 4: /cgn2.6/ptodata/2/1aa/PCUTS.COMB.pep:\*
- 5: /cgn2.6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	52	24.9	199	US-08-211-312-6	Sequence 6, Appl1
2	52	24.9	199	US-08-472-285-6	Sequence 6, Appl1
3	50	23.9	833	US-08-699-103B-4	Sequence 4, Appl1
4	48	23.0	532	US-08-911-321-11	Sequence 11, Appl1
5	48	23.0	532	PCR-US95-13975-2	Sequence 2, Appl1
6	47.5	22.7	283	US-08-852-401-4	Sequence 4, Appl1
7	46	22.0	832	US-08-431-560-1	Sequence 1, Appl1
8	46	22.0	832	US-08-463-345-1	Sequence 1, Appl1
9	46	22.0	1287	US-08-200-232-2	Sequence 1, Appl1
10	46	22.0	1287	PCR-US95-02219-2	Sequence 2, Appl1
11	46	22.0	1287	PCR-US95-02219A-2	Sequence 2, Appl1
12	46	22.0	1296	US-08-470-260-3	Sequence 3, Appl1
13	46	22.0	1296	US-08-471-491-3	Sequence 3, Appl1
14	46	22.0	1296	US-08-466-662-3	Sequence 3, Appl1
15	45.5	21.8	251	US-08-209-747-8	Sequence 8, Appl1
16	45.5	21.8	251	US-08-458-298-8	Sequence 8, Appl1
17	45.5	21.8	331	US-08-356-180-3	Sequence 3, Appl1
18	45.5	21.8	592	US-08-961-083-222	Sequence 222, App
19	45.5	21.8	2233	US-08-569-853-1	Sequence 1, Appl1
20	45.5	21.8	2233	US-08-569-853-2	Sequence 1, Appl1
21	45.5	21.8	2233	US-08-987-439-1	Sequence 1, Appl1
22	45	21.5	462	US-08-865-597A-2	Sequence 2, Appl1
23	45	21.5	626	US-08-956-242-2	Sequence 2, Appl1
24	45	21.5	626	US-09-351-215-2	Sequence 2, Appl1
25	45	21.5	968	US-08-426-236-2	Sequence 2, Appl1
26	44.5	21.3	132	US-08-961-083-194	Sequence 194, App
27	44.5	21.3	359	US-08-457-997B-2	Sequence 2, Appl1
28	44.5	21.3	359	US-08-467-722A-2	Sequence 2, Appl1

29	44	21.1	20	2	US-08-462-221-4	Sequence 4, Appl1
30	44	21.1	32	1	US-08-446-692-21	Sequence 21, Appl1
31	44	21.1	32	2	US-08-488-351A-21	Sequence 21, Appl1
32	44	21.1	74	2	US-08-343-443B-11	Sequence 11, Appl1
33	44	21.1	198	2	US-08-943-915-33	Sequence 33, Appl1
34	44	21.1	207	2	US-08-943-915-32	Sequence 32, Appl1
35	44	21.1	207	2	US-08-943-915-5	Sequence 5, Appl1
36	44	21.1	263	2	US-08-752-844-66	Sequence 66, Appl1
37	44	21.1	339	1	US-08-266-451B-22	Sequence 22, Appl1
38	44	21.1	339	2	US-08-748-725-22	Sequence 22, Appl1
39	44	21.1	340	2	US-08-355-844-1	Sequence 1, Appl1
40	44	21.1	340	4	PCR-US95-16126-1	Sequence 1, Appl1
41	44	21.1	494	3	US-09-198-956-2	Sequence 2, Appl1
42	43.5	20.8	380	2	US-08-472-659-34	Sequence 34, Appl1
43	43.5	20.8	380	2	US-08-474-661-34	Sequence 34, Appl1
44	43.5	20.8	380	2	US-08-611-977-34	Sequence 34, Appl1
45	43.5	20.8	449	2	US-08-657-392-2	Sequence 2, Appl1

## ALIGNMENTS

```
RESULT 1
US-08-211-312-6
; Sequence 6, Application US/08211312
; Patent No. 5986051
; GENERAL INFORMATION:
; APPLICANT: LABIGNE, AGNES
; APPLICANT: CUSAC, VALERIE
; APPLICANT: FERRERO, RICHARD
; TITLE OF INVENTION: GENES OF HELICIOBACTER PYLORI NECESSARY
; TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSEE: P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/211,312
; FILING DATE: 01-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 91 12198
; FILING DATE: 03-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR92/00921
; FILING DATE: 02-OCT-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NO. 5986051man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-075-0XPCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPT UR
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 199 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-211-312-6
```

Query Match 24.9%; Score 52; DB 2; Length 199;  
Best Local Similarity 40.0%; Pred. No. 2.4;  
Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;  
QY 6 NSADAQPYVGA--KIGQVDAKQINKNTAY 33  
DB 145 NKIDAPYVGADLKVMERDSKRIAKSPLF 174

RESULT 2  
US-08-472-285-6  
Sequence 6, Application US/08472285  
Patent No. 6027878  
GENERAL INFORMATION:  
APPLICANT: LABIGNE, AGNES  
APPLICANT: CUSSAC, VALERIE  
APPLICANT: FERRERO, RICHARD  
TITLE OF INVENTION: GENES OF HELICOBACTER PYLORI NECESSARY  
TITLE OF INVENTION: FOR THE REGULATION AND MATURATION OF UREASE AND THEIR USE  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESS: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/472.285  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,312  
FILING DATE: 01-JUL-1994  
APPLICATION NUMBER: FR 91 12198  
FILING DATE: 03-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR92/00921  
FILING DATE: 02-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 6027878man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-075-0XPCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 199 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-472-285-6

Query Match 24.9%; Score 52; DB 3; Length 199;  
Best Local Similarity 40.0%; Pred. No. 2.4;  
Matches 12; Conservative 6; Mismatches 10; Indels 2; Gaps 1;  
QY 6 NSADAQPYVGA--KIGQVDAKQINKNTAY 33  
DB 145 NKIDAPYVGADLKVMERDSKRIAKSPLF 174

RESULT 3  
US-08-699-103B-4

Sequence 4, Application US/08699103B  
Patent No. 6107462  
GENERAL INFORMATION:  
APPLICANT: Rine, Jasper D.  
APPLICANT: Hampton, Randolph  
TITLE OF INVENTION: GENES AND PROTEINS CONTROLLING  
TITLE OF INVENTION: CHOLESTEROL SYNTHESIS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P.C.  
STREET: 2200 Sand Hill Road, Suite 100  
CITY: Menlo Park  
STATE: CA  
COUNTRY: USA  
ZIP: 94025  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/699,103B  
FILING DATE: 16-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/002,581  
FILING DATE: 17-AUG-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D.  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 09272/005001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650/322-5070  
TELEFAX: 650/854-0875  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 833 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-699-103B-4

Query Match 23.9%; Score 50; DB 3; Length 833;  
Best Local Similarity 37.8%; Pred. No. 29;  
Matches 14; Conservative 7; Mismatches 10; Indels 6; Gaps 2;  
QY 6 NSADAQPYV--GAKIGVDAKQINKNTAYGIYAGYN 40  
DB 170 DSAKALLYQRAQOLGNKAKOV----LAYRYSGFN 202

RESULT 4  
US-08-911-321-11  
Sequence 11, Application US/08911321  
Patent No. 6010703  
GENERAL INFORMATION:  
APPLICANT: Roger K. Maes and Stephen J. Spatz  
TITLE OF INVENTION: Recombinant Poxvirus  
TITLE OF INVENTION: Vaccine Against  
TITLE OF INVENTION: Peline Rhinotracheitis  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ian C. McLeod  
STREET: 2190 Commons Parkway  
CITY: Okemos  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911.321  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/096.183  
FILING DATE: July 26, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-166  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: NO. 60107036  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532  
TYPE: Amino Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE:  
DESCRIPTION: Polypeptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Feline herpesvirus-1  
STRAIN: 1  
INDIVIDUAL ISOLATE: C-27  
CELL TYPE: N/A  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: Deduced Sequence  
OTHER INFORMATION: 9E  
US-08-911-321-11

Query Match 23.0%; Score 48; DB 3; Length 532;  
Best Local Similarity 36.7%; Pred No. 33;  
Matches 11; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

Qy 7 SADAOPYGAKIGQVDAKQNGKNTAYGIY 36  
1 : 11 : 11 1 1 1 1 1 1  
Db 299 SINHMPYIEQPNNVDLKFINVPTNASGLY 328

RESULT 5  
PCT-US95-13975-2  
Sequence 2, Application PC/TUS9513975  
GENERAL INFORMATION:  
APPLICANT: Cochran, Mark D.  
APPLICANT: McDonnell, Michael W.  
TITLE OF INVENTION: Recombinant Feline Herpes Virus  
NUMBER OF SEQUENCES: 73  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: John P. White  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM 330 466 DX2  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/13975  
FILING DATE: 26-OCT-1995  
CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/329,883  
FILING DATE: 26-OCT-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: White, John P.  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 39118-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)278-0400  
TELEFAX: (212)391-0525  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 532 amino acids  
TYPE: amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: protein  
PCT-US95-13975-2

Query Match 23.0%; Score 48; DB 4; Length 532;  
Best Local Similarity 36.7%; Pred No. 33;  
Matches 11; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

Qy 7 SADAOPYGAKIGQVDAKQNGKNTAYGIY 36  
1 : 11 : 11 1 1 1 1 1 1  
Db 299 SINHMPYIEQPNNVDLKFINVPTNASGLY 328

RESULT 6  
US-08-852-401-4  
Sequence 4, Application US/08852401  
Patent No. 5976836  
GENERAL INFORMATION:  
APPLICANT: Weber, J. Mark  
APPLICANT: Hessler, Paul E.  
APPLICANT: Larsen, Peter E.  
APPLICANT: Liu, B. Minh  
TITLE OF INVENTION: Methods and Compositions for Enhancing  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rocky, Milnamow & Katz, Ltd.  
STREET: 2 Prudential Plaza, 180 N. Stetson, Suite  
STREET: 4700  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/852,401  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa L.  
REGISTRATION NUMBER: 38,978  
REFERENCE/DOCKET NUMBER: FER2159POO300US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 283 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-852-401-4





```

: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (510) 601-2708
: TELEFAX: (510) 655-3542
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1296 amino acids
:   TYPE: amino acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-470-260-3

Query Match                22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db 1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 13
: Sequence 3, Application US/08471491B
: Patent No. 6090611
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: APPLICANT: Bugnoli, Massimo
: APPLICANT: Telford, John
: APPLICANT: Macchia, Giovanni
: APPLICANT: Rappuoli, Rino
: TITLE OF INVENTION: Helicobacter pylori Proteins Useful For Vaccines And
: FILE REFERENCE: CHIR0044
: CURRENT APPLICATION NUMBER: US/08/471,491B
: CURRENT FILING DATE: 1995-06-06
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 1296
: TYPE: PRT
: ORGANISM: Helicobacter pylori
: US-08-471-491-3

Query Match                22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db 1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 14
: Sequence 3, Application US/0846662B
: Patent No. 6130059
: GENERAL INFORMATION:
: APPLICANT: Covacci, Antonello
: APPLICANT: Bugnoli, Massimo
: APPLICANT: Telford, John
: APPLICANT: Macchia, Giovanni
: APPLICANT: Rappuoli, Rino
: TITLE OF INVENTION: Helicobacter pylori Proteins Useful For Vaccines And
: FILE REFERENCE: CHIR0057
: CURRENT APPLICATION NUMBER: US/08/466,662B
: CURRENT FILING DATE: 1995-06-06
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
```

```

: LENGTH: 1296
: TYPE: PRT
: ORGANISM: Helicobacter pylori
: US-08-466-662-3

Query Match                22.0%; Score 46; DB 3; Length 1296;
Best Local Similarity 40.5%; Pred. No. 2e+02;
Matches 15; Conservative 5; Mismatches 11; Indels 6; Gaps 3;

OY 4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db 1049 YGTSAGVDAYLN---GQVEA--IVGGFGSYG-YSSFN 1079

RESULT 15
: Sequence 8, Application US/08209747
: Patent No. 5733771
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: Colgin, Mark
: TITLE OF INVENTION: CDNA's Encoding Minor Ampullate Spider
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: APPLICATION DATA:
: APPLICATION NUMBER: US/08/209,747
: FILING DATE: 14-MAR-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 251 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-209-747-8

Query Match                21.8%; Score 45.5; DB 1; Length 251;
Best Local Similarity 35.1%; Pred. No. 30;
Matches 13; Conservative 3; Mismatches 12; Indels 9; Gaps 2;

OY 4 YGNSADAQPYVGAKIGVDAKQINGKNTAVGIYAGYN 40
Db 90 YGCGAGS---GAR-----AASAGAGTGYGSSGCGYN 117

Search completed: March 6, 2001, 14:37:04
Job time: 147 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:35:47 ; Search time 23.41 Seconds

(without alignments)  
116.020 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209

Sequence: 1 A1SYGNSADAQPYGAKIGQVDAKQJNGKNTAYGYAGYN 40

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR\_66:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	59.5	28.5	380	2 T24786	hypothetical prote
2	59	28.2	160	2 H82207	hypothetical prote
3	55.5	26.6	289	2 S38806	potin opma - Rhodo
4	55.5	26.6	1196	2 S40908	TyB protein - yeas
5	55.5	26.6	1278	2 B22671	TyB protein - yeas
6	55.5	26.6	1328	2 S52481	TyB protein - yeas
7	55.5	26.6	1328	2 S52894	TyB protein - yeas
8	55.5	26.6	1328	2 S50947	TyB protein - yeas
9	55.5	26.6	1328	2 B28097	TyB protein - yeas
10	55.5	26.6	1348	2 B23496	TyB protein - yeas
11	55.5	26.6	1749	2 S69972	TyB protein - yeas
12	55.5	26.6	1755	2 S50641	TyB protein YER138
13	55.5	26.6	1755	2 S50663	TyB protein - yeas
14	55.5	26.6	1755	2 S57047	TyB protein - yeas
15	55.5	26.6	1755	2 S57045	TyB protein - yeas
16	55.5	26.6	1755	2 S45736	TyB protein - yeas
17	55.5	26.6	1755	2 S69963	TyB protein - yeas
18	55.5	26.6	1755	2 S69839	TyB protein - yeas
19	55.5	26.6	1755	2 S69980	TyB protein - yeas
20	55.5	26.6	1755	2 S69957	TyB protein - yeas
21	55.5	26.6	1755	2 S69951	TyB protein - yeas
22	55.5	26.6	1755	2 S69969	TyB protein - yeas
23	55.5	26.6	1755	2 S69979	TyB protein - yeas
24	55.5	26.6	1755	2 S69979	TyB protein - yeas
25	55.5	26.6	1755	2 S69838	TyB protein - yeas
26	55.5	26.6	1755	2 S70298	TyB protein - yeas
27	55.5	26.6	1755	2 S69955	TyB protein - yeas
28	55.5	26.6	1755	2 S69845	TyB protein - yeas
29	55.5	26.6	1755	2 S69866	TyB protein - yeas

30	55.5	26.6	1755	2 S69982	TyB protein - yeas
31	55.5	26.6	1755	2 S69949	TyB protein - yeas
32	55.5	26.6	1755	2 S61763	TyB protein - yeas
33	55.5	26.6	1755	2 S69975	TyB protein - yeas
34	55.5	26.6	1755	2 S45867	TyB protein - yeas
35	55.5	26.6	1756	2 S69983	TyB protein - yeas
36	55.5	26.6	1770	2 S45842	TyB protein - yeas
37	55.5	26.6	1770	2 S69953	TyB protein - yeas
38	55.5	26.6	1770	2 S69948	TyB protein - yeas
39	55.5	26.6	1770	2 S70233	TyB protein - yeas
40	55.5	26.6	1770	2 S70230	TyB protein - yeas
41	55.5	26.6	1770	2 S69966	TyB protein - yeas
42	55.5	26.6	1770	2 S69950	TyB protein - yeas
43	55.5	26.6	1770	2 S58651	TyB protein - yeas
44	55.5	26.6	1771	2 S53592	TyB protein - yeas
45	55.5	26.6	1793	2 S52601	TyB protein - yeas

## ALIGNMENTS

```
RESULT 1
T24786
Hypothetical protein ZK1321.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T24786; T27760
R:Gardner, A.
Submitted to the EMBL Data Library, March 1995
A:Reference number: Z19935
A:Accession: T24786
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <MIT>
A:Cross-references: EMBL:Z48717; PIDN:CA88612.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone T10B9
R:Gardner, A.
Submitted to the EMBL Data Library, March 1995
A:Reference number: Z20415
A:Accession: T27760
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-380 <MI2>
A:Cross-references: EMBL:Z48584; PIDN:CA88478.1; GSPDB:GN00020; CESP:ZK1321.4
A:Experimental source: clone ZK1321
C:Genetics:
A:Gene: CESP:ZK1321.4
A:Map position: 2
A:Introns: 4/3; 35/2; 85/3; 115/2; 224/2; 328/2; 346/3
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1321.4

Query Match          28.5%; Score 59.5; DB 2; Length 380;
Best Local Similarity 38.6%; Pred. No. 3.1;
Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY 3 SYGNSADAQPYGAKIGQVDAKQIN-----GKTFAYGIAG 38
DB 224 STNNGATSQPY-GAGSGGTPLNOMFTINTSPAPGANGAYGAOCG 266

RESULT 2
H82207
Hypothetical protein VC1384 [imported] - Vibrio cholerae (group O1 strain N16961)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Drygoli, I.; Sellers
  L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
```

```

A:Accession: H82207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-160 <HEI>
A:Cross-references: GB:AE004217; GB:AE003852; NID:9655866; PIDN:AAE94542.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16561; biotype E1 Tor
C:Genetics:
A:Gene: VC1384
A:Map position: 1

Query Match      28.2%  Score 59;  DB 2;  Length 160;
Best Local Similarity 45.2%;  Pred. No. 1.5;
Matches 14;  Conservative 6;  Mismatches 9;  Indels 2;  Gaps 1;

Oy      8  ADAOPYGAKIGOVDAKOINGKNTAYGIGY 38
      11: 111111111111111111111111
Db      19  ADSWITGASVGSDEYKKG--TAYSVHAG 47

RESULT 3
S38806
porin opma - Rhodopseudomonas blastica
C:Species: Rhodopseudomonas blastica
C:Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 07-May-1999
A:Accession: S38806; S38793
R:Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schult, G.E.
Protein Sci. 3, 58-63, 1994
A:Title: Structure of the membrane channel porin from Rhodopseudomonas blastica at 2.0 Å
A:Reference number: S38806; MUID:94191532
A:Accession: S38806
A:Status: nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-289 <KR>
R:Kreusch, A.; Neubueser, A.; Schiltz, E.; Weckesser, J.; Schult, G.E.
Submitted to the Protein Sequence Database, November 1993
A:Reference number: S38793
A:Accession: S38793
A:Molecule type: protein
A:Residues: 1-71, 'X', 73-82; 91-104; 109-137; 146-163; 173-289 <KR2>
C:Genetics:
A:Gene: opma
C:Keywords: homotrimer; membrane protein

Query Match      26.6%  Score 55.5;  DB 2;  Length 289;
Best Local Similarity 41.7%;  Pred. No. 8;
Matches 15;  Conservative 3;  Mismatches 15;  Indels 3;  Gaps 1;

Oy      4  YGNSADAPYVCAKIGOVDAKOINGKNTAYGIGY 39
      111111111111111111111111111111
Db      227  YGNVAFGATTVARAYVSDIDRA---GADTAYGIGADY 259

RESULT 4
S40908
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.A
N:Alternate names: protein YAR009C
C:Species: Saccharomyces cerevisiae
C:Date: 13-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 22-Oct-1999
A:Accession: S40908
R:Clark, M.W.; Keng, T.; Storms, R.K.; Zhong, W.; Fortin, N.; Zeng, B.; Delaney, S.; Quong,
Submitted to the EMBL Data Library, November 1993
A:Description: Sequencing of chromosome I of Saccharomyces cerevisiae: Analysis of the 4
A:Reference number: S40891
A:Accession: S40908
A:Molecule type: DNA
A:Residues: 1-1196 <CLAG>
A:Cross-references: EMBL:L22015; NID:g1339990; PIDN:AMC04967.1; PID:g2564964
C:Genetics:
A:Map position: 1R
A:Mobile element: retrotransposon Ty1.A
A:Superfamily: Tyb protein

```

```

Query Match Similarity 26.6% Score 55.5 DB 2 Length 1196;
Best Local Similarity 43.8% Pred. No. 33;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVGAKIGOV---DAKOINGKMT 31
    |||| ||::||: |||||
DB 1053 SYGN---QPYKKSQIGNIYLNGKVGIGKST 1080

RESULT 5
B22671
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty912
C:Species: Saccharomyces cerevisiae
C:Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 26-Aug-1999
C:Accession: B22671; S05890
R:Clare, J.; Farabaugh, P.
Proc. Natl. Acad. Sci. U.S.A. 82, 2829-2833, 1985
A>Title: Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism
A:Reference number: A94039; MUID:85190589
A:Accession: B22671
A:Molecule type: DNA
A:Residues: 1-1278 <CIA>
R:Mellor, J.; Fulton, S.M.; Dobson, M.J.; Wilson, W.; Kingsman, S.M.; Kingsman, A.J.
Nature 313, 243-246, 1985
A>Title: A retrovirus-like strategy for expression of a fusion protein encoded by yea
A:Reference number: A21856; MUID:85111117
A:Accession: S05890
A:Molecule type: DNA
A:Residues: 1-6,'1',8-10,'K',12-53 <MEL>
A:Cross-references: EMBL:X01736; NID:94708; PIDN:CAA2874.1; PID:g1326020
C:Genetics:
A:Mobile element: retrotransposon Ty912
C:Superfamily: Tyb protein

Query Match Similarity 26.6% Score 55.5 DB 2 Length 1278;
Best Local Similarity 43.8% Pred. No. 35;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVGAKIGOV---DAKOINGKMT 31
    |||| ||::||: |||||
DB 1135 SYGN---QPYKKSQIGNIYLNGKVGIGKST 1162

RESULT 6
S52481
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1
N:Alternate names: protein YML039w
C:Species: Saccharomyces cerevisiae
C:Date: 08-May-1995 #sequence_revision 21-Jul-1995 #text_change 20-Jun-2000
C:Accession: S52481
R:Connor, R.; Churcher, C.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52478
A:Accession: S52481
A:Molecule type: DNA
A:Residues: 1-1328 <CON>
A:Cross-references: EMBL:Z48430; NID:g683664; PIDN:CAA8330.1; PID:g1326005
C:Genetics:
A:Map position: 13L
A:Mobile element: retrotransposon Ty1
C:Superfamily: Tyb protein

Query Match Similarity 26.6% Score 55.5 DB 2 Length 1328;
Best Local Similarity 43.8% Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAQPYVGAKIGOV---DAKOINGKMT 31
    |||| ||::||: |||||
DB 1185 SYGN---QPYKKSQIGNIYLNGKVGIGKST 1212

```



```
RESULT 7
SY2894
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-9532
N:Alternate names: protein YM9532.10c; protein YMR045c
C:Species: Saccharomyces cerevisiae
C>Date: 08-May-1995 #sequence_revision 01-Sep-1995 #text_change 20-Jun-2000
C:Accession: S52894
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, February 1995
A:Reference number: S52885
A:Accession: S52894
A:Molecule type: DNA
A:Residues: 1-1328 <ODE>
A:Cross-references: EMBL:Z48502; NID:g695715; PIDN:CAA88411.1; PID:g1326011
C:Genetics:
A:Map position: 13R
A:Mobile element: retrotransposon Ty1-9532
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN----QPYKSQLGNITFLNGKVIIGKST 1212

RESULT 8
SY0948
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1
N:Alternate names: protein YM9827.08; protein YML045w
C:Species: Saccharomyces cerevisiae
C>Date: 10-Feb-1995 #sequence_revision 12-May-1995 #text_change 20-Jun-2000
C:Accession: S50948
R:Odell, C.; Bowman, S.
submitted to the EMBL Data Library, January 1995
A:Reference number: S50941
A:Accession: S50948
A:Molecule type: DNA
A:Residues: 1-1328 <ODE>
A:Cross-references: EMBL:Z47816; NID:g642303; PIDN:CAA87830.1; PID:g1326015
C:Genetics:
A:Map position: 13L
A:Mobile element: retrotransposon Ty1
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN----QPYKSQLGNITFLNGKVIIGKST 1212

RESULT 9
B28097
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-H3 (fragment)
C:Species: Saccharomyces cerevisiae
C>Date: 03-Nov-1988 #sequence_revision 25-Apr-1997 #text_change 25-Apr-1997
C:Accession: B28097
R:Boeke, J.D.; Eichinger, D.; Castrillon, D.; Fink, G.R.
Mol. Cell. Biol. 8, 1432-1442, 1988
A:Title: The Saccharomyces cerevisiae genome contains functional and nonfunctional copies
A:Reference number: A28097; MUID:88246410
A:Accession: B28097
A:Molecule type: DNA
A:Residues: 1-1328 <BOE>
```

```
A:Cross-references: EMBL:M18706
C:Genetics:
A:Mobile element: retrotransposon Ty1-H3
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1328;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1185 SYGN----QPYKSQLGNITFLNGKVIIGKST 1212

RESULT 10
B23496
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1-17
C:Species: Saccharomyces cerevisiae
C>Date: 20-Aug-1987 #sequence_revision 20-Aug-1987 #text_change 25-Apr-1997
C:Accession: B23496
R:Warrington, J.R.; Waring, R.B.; Newlon, C.S.; Indge, K.J.; Oliver, S.G.
Nucleic Acids Res. 13, 6679-6693, 1985
A:Title: Nucleotide sequence characterization of Ty 1-17, a class II transposon from
A:Reference number: A93591; MUID:86041864
A:Accession: B23496
A:Molecule type: DNA
A:Residues: 1-1348 <MAR>
A:Cross-references: EMBL:X03840; EMBL:J01333; EMBL:K00633; EMBL:X02991
C:Genetics:
A:Mobile element: retrotransposon Ty1-17
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1348;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1205 SYGN----QPYKSQLGNITFLNGKVIIGKST 1232

RESULT 11
S69972
TYB protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.N
N:Alternate names: protein N2453
C:Species: Saccharomyces cerevisiae
C>Date: 09-Mar-1996 #sequence_revision 06-Sep-1996 #text_change 20-Jun-2000
C:Accession: S69972
R:Bergez, P.; Dolgon, F.; Crouzet, M.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62975
A:Accession: S69972
A:Molecule type: DNA
A:Residues: 1-1749 <BER>
A:Cross-references: EMBL:Z71330; NID:g1301918; PIDN:CAA95924.1; PID:g1301920
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
C:Genetics:
A:Map position: 14L
A:Mobile element: retrotransposon Ty1.N
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1749;
Best Local Similarity 43.8%; Pred. No. 48;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAPYVGAKIGQV---DAKQINGKNT 31
||||| ||| :||| : |||||
Db 1606 SYGN----QPYKSQLGNITFLNGKVIIGKST 1633
```

```
RESULT 12
S50641
Tyb protein YER138c - yeast (Saccharomyces cerevisiae) retrotransposon Ty
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999
C:Accession: S50641
R:Diehl, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50428
A:Accession: S50641
A:Molecule type: DNA
A:Residues: 1-1755 <DIE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64665.1; PID:g603378; MIPS:YER138c
C:Genetics:
A:Map position: 5R
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKIVIGKST 1639

RESULT 13
S50663
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.E
N:Alternate names: protein YER160C
C:Species: Saccharomyces cerevisiae
C:Date: 28-May-1993 #sequence_revision 24-Feb-1995 #text_change 26-Aug-1999
C:Accession: S50663; S30812; S53556
R:Diehl, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 8229, 9115, 9132, 9981, and lambda
A:Reference number: S50428
A:Accession: S50663
A:Molecule type: DNA
A:Residues: 1-1755 <DIE>
A:Cross-references: EMBL:U18917; NID:g603377; PIDN:AAB64687.1; PID:g603400
R:Mulligan, J.T.; Diehl, F.S.; Hennessy, K.M.; Sehl, P.; Komp, C.; Wei, Y.; Taylor,
submitted to the EMBL Data Library, February 1993
A:Reference number: S30812
A:Accession: S30812
A:Molecule type: DNA
A:Residues: 1-230 <MUL>
A:Cross-references: EMBL:L10718
C:Genetics:
A:Map position: 5R
A:Mobile element: retrotransposon Ty1.E
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKIVIGKST 1639

RESULT 14
S57047
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein J1570; protein YUR029w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
C:Accession: S57047; S60512
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; Herba
```

```
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57040
A:Accession: S57047
A:Molecule type: DNA
A:Residues: 1-1755 <HUA>
A:Cross-references: EMBL:Z49528
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading
A:Reference number: S60503; MUID:96109930
A:Accession: S60512
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'T1STFT', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect readi
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:|:| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKIVIGKST 1639
```

```
RESULT 15
S57045
Tyb protein - yeast (Saccharomyces cerevisiae) retrotransposon Ty1.JR
N:Alternate names: protein J1560; protein YUR027w
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change 17-Mar-1999
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
submitted to the Protein Sequence Database, September 1995
A:Reference number: S57040
A:Accession: S57045
A:Molecule type: DNA
A:Residues: 1-1755 <HUA>
A:Cross-references: EMBL:Z49526
A:Note: biosynthesis of this protein involves a +1 frameshift in the codon for residu
R:Zagulski, M.; Babinska, B.; Gromadka, R.; Migdalski, A.; Rytko, J.; Sulicka, J.; He
A:Title: The sequence of 24.3 kb from chromosome X reveals five complete open reading
A:Reference number: S60503; MUID:96109930
A:Accession: S60511
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'T1STFT', 435-1755 <ZAG>
A:Cross-references: EMBL:X87297
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995
A:Note: the difference at the amino end is due to translation from an incorrect readi
C:Genetics:
A:Map position: 10R
A:Mobile element: retrotransposon Ty1.JR
C:Superfamily: Tyb protein

Query Match          26.6%; Score 55.5; DB 2; Length 1755;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

Oy      3 SYGNSADAQPYVGAKIGOV---DAKQINGKNT 31
      |||| |||::||:~||| |||||
Db      1612 SYGN-----QPYKKSQIGNITLNGKIVIGKST 1639
```

Tue Mar 6 16:25:07 2001

us-09-164-714-1.rpr

Page 5

Search completed: March 6, 2001, 14:37:44  
Job time: 117 sec

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:37:12 ; Search time 14.8 Seconds  
(without alignments)  
87,281 Million cell updates/sec

Title: US-09-164-714-1

Perfect score: 209  
Sequence: 1 A1SYGNSADAQPYVGAKIGQVDAKQINGKNTAYGIVAGYN 40

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	55.5	26.6	289	1	P39767 rhodopseudo
2	55.5	26.6	1328	1	Q03434 saccharomyc
3	55.5	26.6	1328	1	Q04711 saccharomyc
4	55.5	26.6	1328	1	Q04214 saccharomyc
5	55.5	26.6	1328	1	Q04670 saccharomyc
6	55.5	26.6	1347	1	P25384 saccharomyc
7	55.5	26.6	1755	1	P47098 saccharomyc
8	55.5	26.6	1755	1	P47100 saccharomyc
9	54	25.8	1092	1	P54358 diosophilla
10	53.5	25.6	493	1	Q05807 mycobacteri
11	53	25.4	1249	1	Q64560 rattus norv
12	51	24.4	199	1	Q94227 helicobacte
13	51	24.4	199	1	Q09066 helicobacte
14	50.5	24.2	250	1	P72077 neisseria g
15	50	23.9	121	1	P45088 haemophilus
16	50	23.9	325	1	P39261 bacterioph
17	49.5	23.7	77	1	Q74966 schizosacch
18	49.5	23.7	449	1	P34698 streptococc
19	49.5	23.7	584	1	P35077 bordelella
20	49	23.4	357	1	P29356 spinacia ol
21	49	23.4	358	1	P22197 arabidopsis
22	49	23.4	540	1	P25026 rhizopus ol
23	49	23.4	959	1	Q13914 schizosacch
24	49	23.4	1116	1	P38518 bacillus br
25	49	23.4	1262	1	Q64514 mus musculu
26	48.5	23.2	453	1	Q01134 rattus norv
27	48.5	23.2	1156	1	Q09031 bacillus th
28	48	23.0	131	1	P54067 mechanococc
29	48	23.0	204	1	Q55057 streptococc
30	48	23.0	206	1	P72955 synecocyst
31	48	23.0	313	1	P15703 saccharomyc
32	48	23.0	518	1	P56959 mus musculu
33	48	23.0	1088	1	Q48901 glycine max

34	48	23.0	1242	1	P02487 african swi
35	48	23.0	1842	1	Q10289 schizosacch
36	47.5	22.7	472	1	P21948 escherichia
37	47.5	22.7	629	1	P17382 human papil
38	47.5	22.7	1133	1	P36775 saccharomyc
39	47	22.5	241	1	P28034 petroselinu
40	47	22.5	242	1	Q43237 vitis vinif
41	47	22.5	351	1	P40990 saccharomyc
42	47	22.5	360	1	Q9x59 sus scrofa
43	47	22.5	750	1	P34231 saccharomyc
44	47	22.5	788	1	P37254 saccharomyc
45	46.5	22.2	329	1	P43819 haemophilus

## ALIGNMENTS

RESULT	ID	PORT_RHOBL	STANDARD	PRT	289 AA.
AC	P39767				
DT	01-FEB-1995	(Rel. 31, Created)			
DT	01-FEB-1995	(Rel. 31, Last sequence update)			
DT	15-JUL-1999	(Rel. 38, Last annotation update)			
DE	PORIN.				
GN	OPMA.				
OS	Rhodopseudomonas blattica.				
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.				
RN	[1]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).				
RX	MEDLINE=94191532; PubMed=8142898;				
RA	Kreusch A., Neubauer A., Schiltz E., Weckesser J., Schulz G.E.;				
RT	"Structure of the membrane channel porin from Rhodopseudomonas blattica at 2.0-A resolution."				
RL	Protein Sci. 3:58-63(1994).				
RN	[2]				
RP	X-RAY CRYSTALLOGRAPHY (1.96 ANGSTROMS).				
RX	MEDLINE=95055730; PubMed=7525973;				
RA	Kreusch A., Schulz G.E.;				
RT	"Refined structure of the porin from Rhodopseudomonas blattica. Comparison with the porin from Rhodobacter capsulatus."				
RL	J. Mol. Biol. 243:891-905(1994).				
RN	[3]				
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF MUTANTS.				
RA	Schmid B., Weveraud L., Kromer M., Schulz G.E.;				
RT	"Porin mutants with new channel properties."				
RL	Protein Sci. 7:1603-1611(1998).				
CC	-1- FUNCTION: FORMS CHANNELS THAT ALLOW THE PASSIVE DIFFUSION OF SMALL HYDROPHILIC SOLUTES UP TO AN EXCLUSION LIMIT OF ABOUT 0.6 KDA.				
CC	-1- SUBUNIT: HOMOTRIMER.				
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.				
DR	PIR: S38806; S38806.				
DR	PDB: 1PRN; 24-OCT-94.				
DR	PDB: 2PRN; 13-JAN-99.				
DR	PDB: 3PRN; 12-AUG-98.				
DR	PDB: 5PRN; 12-AUG-98.				
DR	PDB: 6PRN; 12-AUG-98.				
DR	PDB: 7PRN; 12-AUG-98.				
DR	PDB: 8PRN; 12-AUG-98.				
DR	PDB: 1BH3; 12-AUG-98.				
KW	Outer membrane; Transmembrane; Porin; 3D-structure.				
SEQ	SEQUENCE 289 AA; 30597 MW; 0835209803A1044C CRC64;				
Query Match	26.6%; Score 55.5; DB 1; Length 289;				
Best Local Similarity	41.7%; Pred. No. 3.7;				
Matches	15; Conservative 3; Mismatches 15; Indels 3; Gaps 1;				
QY	4 YGNSADAPYVGAKIGQVDAKQINGKNTAYGIVAGY 39				
DB	227 YGNYAFGATTVRAVYSDIDRA--GADTAYGIGADY 259				

```
RESULT 2
YMD9_YEAST STANDARD: PRT; 1328 AA.
AC 003434;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONON TY1 PROTEIN B.
FT TY1B OR YML039W OR YM8054.04.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Connor R., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 248430; CAAB8330.1; -.
DR SGD; S0004503; YML039W.
DR INTERPRO; IPR001969; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KW Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 151036 MW; C7D14E1AA675E93 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSSQIGNITLNGKVGIGKST 1212

RESULT 3
YME4_YEAST STANDARD: PRT; 1328 AA.
AC 004711;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONON TY1 PROTEIN B.
FT TY1B OR YML044W OR YM9827.08.
GN Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 248502; CAAB8411.1; -.
DR SGD; S0004648; YMR045C.
DR INTERPRO; IPR001969; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KW Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 151187 MW; AA19E50B62BA43F95 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSSQIGNITLNGKVGIGKST 1212

RESULT 5
YMU0_YEAST STANDARD: PRT; 1328 AA.
AC 004670;
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 247816; CAAB7830.1; -.
DR INTERPRO; IPR001969; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KW Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 150945 MW; E9B964CD76CD281F CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSSQIGNITLNGKVGIGKST 1212

RESULT 4
YMT5_YEAST STANDARD: PRT; 1328 AA.
AC 004214;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSPONON TY1 PROTEIN B.
GN TY1B OR YMR045C OR YM9532.10C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Odell C., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 248502; CAAB8411.1; -.
DR SGD; S0004648; YMR045C.
DR INTERPRO; IPR001969; -.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
KW Transposable element; Hypothetical protein; Hydrolase;
KW Aspartyl protease; ATP-binding.
FT ACT_SITE 34 34 PROTEASE (BY SIMILARITY).
FT NP_BIND 1204 1211 ATP (POTENTIAL).
SQ SEQUENCE 1328 AA; 151187 MW; AA19E50B62BA43F95 CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;
Best Local Similarity 43.8%; Pred. No. 17;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASDAQPYVGAKIGOV---DAKQINGKNT 31
Db 1185 SYGN---QPYKSSQIGNITLNGKVGIGKST 1212

RESULT 5
YMU0_YEAST STANDARD: PRT; 1328 AA.
AC 004670;
```

DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSDON TY1 PROTEIN B.  
 GN TY1B OR YMR050C OR YMR9796.03C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-S288C / AB972.  
 RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A.;  
 RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 -----  
 CC EMBL: Z49703; CAA89760.1; -  
 CC SGD: S0004653; YMR050C.  
 DR INTERPRO: IPR001969; -  
 DR PROSITE: PS00141; ASP\_PROTEASE: 1.  
 KW Transposable element; Hypothetical protein; Hydrolase;  
 KM Aspartyl protease; ATP-binding  
 FT ACT\_SITE 34 34 PROTEASE (BY SIMILARITY).  
 FT NP\_BIND 1204 1211 ATP (POTENTIAL).  
 FT SEQUENCE 1328 AA; 150967 MW; DACD7A471697DIDD CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1328;  
 Best Local Similarity 43.8%; Pred. No. 17;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADQPYGAKIGOV---DAKQINGKNT 31  
 |||| ||| :||| : | |||:  
 DB 1185 SYGN---QPYKSSQIGNIFLNGKVIQGST 1212

RESULT 6  
 YC89\_YEAST STANDARD; PRT; 1347 AA.  
 AC P25384;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE TRANSDON TY1-17 PROTEIN B.  
 GN TY1B OR YCL019W OR YCL19W.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-86205247; PubMed-3010239;  
 RA Warrington J.R., Anwar R., Newton C.S., Waring R.B., Davies R.W.,  
 RA Indge K.J., Oliver S.G.;  
 RA "A 'hot-spot' for Ty transposition on the left arm of yeast  
 RT chromosome III.";  
 RL Nucleic Acids Res. 14:3475-3485(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-86041864; PubMed-2997719;  
 RA Warrington J.R., Waring R.B., Newton C.S., Indge K.J., Oliver S.G.;  
 RT "Nucleotide sequence characterization of Ty 1-17, a class II  
 RT transposon from yeast.";  
 RL Nucleic Acids Res. 13:6679-6693(1985).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,

RA Stateva L.I.;  
 RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 -----  
 CC EMBL: X03840; CAA27458.1; -  
 CC EMBL: X59720; E264443; -  
 DR PIR: B23496; B23496.  
 DR PIR: S19345; S19345.  
 DR MEROPS: A11.003; -  
 DR SGD: S0000524; YCL019W.  
 KW Transposable element; Hypothetical protein; Hydrolase;  
 KM Aspartyl protease; ATP-binding.  
 FT NP\_BIND 1223 1230 ATP (POTENTIAL).  
 FT SEQUENCE 1347 AA; 154069 MW; AD3660C5E7B282NF CRC64;

Query Match 26.6%; Score 55.5; DB 1; Length 1347;  
 Best Local Similarity 43.8%; Pred. No. 17;

Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNSADQPYGAKIGOV---DAKQINGKNT 31  
 |||| ||| :||| : | |||:  
 DB 1204 SYGN---QPYKSSQIGNIFLNGKVIQGST 1231

RESULT 7  
 YJZ7\_YEAST STANDARD; PRT; 1755 AA.  
 AC P47098; P47194;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSDON TY1 PROTEIN B.  
 GN TY1B OR YJR027W OR J1560.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;  
 CC Saccharomycetaceae; Saccharomyces.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-9610930; PubMed-8619316;  
 RA Zaglinski M., Babinaka B., Gromadka R., Migdalski A., Rylka J.,  
 RA Sulicka J., Herbert C.J.;  
 RA "The sequence of 24.3 kb from chromosome X reveals five complete open  
 RT reading frames, all of which correspond to new genes, and a tandem  
 RT insertion of a Ty1 transposon.";  
 RL Yeast 11:1179-1186(1995).  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
 -----  
 CC EMBL: Z49526; CAA89553.1; -  
 CC MEROPS: A11.003; -  
 DR SGD: S0003788; YJR027W.  
 DR INTERPRO: IPR001042; -  
 DR PFAM: PF01021; TYA: 1.  
 DR PROSITE: PS00141; ASP\_PROTEASE: 1.  
 KW Transposable element; Hypothetical protein; Hydrolase;  
 KM Aspartyl protease; ATP-binding.  
 FT ACT\_SITE 461 461 PROTEASE (BY SIMILARITY).

```

ET  NP_HIND 1631 1638 ATP (POTENTIAL).
SQ  SEQUENCE 1755 AA; 198615 MW; 04E248A7FD7596F CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 1; Length 1755;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASADAPYVGAKIGOV---DAKOINGKNT 31
    |||||  |||  ::|||  :  |||  |||
Db  1612 SYGN---OPYYKSOIGNIFLNGKRVIGGKST 1639

RESULT 8
YJ29_YEAST
ID  YJ29_YEAST STANDARD: PRT: 1755 AA.
AC  P47100; P87195;
DT  01-FEB-1996 (Rel. 33, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DE  30-MAY-2000 (Rel. 39, Last annotation update)
DE  TRANSPOSON T11 PROTEIN B.
DE  TY1B OR YJR029W OR J1570.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC  Saccharomycetaceae; Saccharomyces.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE-96109930; PubMed-8619316;
RX  Zaglinski M., Babinska B., Gromadka R., Migdalski A., Rytko J.,
RX  Sulicka J., Herbert C.J.;
RT  "The sequence of 24.3 kb from chromosome X reveals five complete open
RT  reading frames, all of which correspond to new genes, and a tandem
RT  insertion of a ty1 transposon."
RL  Yeast 11:1179-1186(1995).
CC  -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY U23.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch)
-----
DR  EMBL; Z49528; CAA89556.1; -.
DR  MEROPS; A11.003; -.
DR  SCD; S0003790; YJR029W.
DR  INTERPRO: IPR001042; -.
DR  PFAM: PF01021; TYA: 1.
DR  PROSITE: PS00141; ASP_PROPEASE; 1.
KW  Transposable element; Hypothetical protein; Hydrolase;
KW  Aspartyl protease; ATP-binding.
FT  ACT_SITE 461 461 PROPEASE (BY SIMILARITY).
FT  NP_BIND 1631 1638 ATP (POTENTIAL).
SQ  SEQUENCE 1755 AA; 198592 MW; C8B125BF2967C990 CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 1; Length 1755;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

OY 3 SYGNASADAPYVGAKIGOV---DAKOINGKNT 31
    |||||  |||  ::|||  :  |||  |||
Db  1612 SYGN---OPYYKSOIGNIFLNGKRVIGGKST 1639

RESULT 9
DPOD_DROME
ID  DPOD_DROME STANDARD: PRT: 1092 AA.
AC  P54358;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)

```

```

DE  DNA POLYMERASE DELTA CATALYTIC CHAIN (EC 2.7.7.7).
GN  POLD.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscophila;
OC  Ephydroidea; Drosophilidae; Drosophila.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE-96125196; PubMed-8543168;
RX  Chiang C.S., Lehman I.R.;
RT  "Isolation and sequence determination of the cDNA encoding DNA
RT  polymerase delta from Drosophila melanogaster."
RL  Gene 166:237-242(1995).
CC  -1- FUNCTION: POSSESSES TWO ENZYMATIC ACTIVITIES: DNA SYNTHESIS
CC  (POLYMERASE) AND AN EXONUCLEOTIC ACTIVITY THAT DEGRADES SINGLE
CC  STRANDED DNA IN THE 3' TO 5' DIRECTION. REQUIRED WITH ITS
CC  ACCESSORY PROTEINS (PROLIFERATING CELL NUCLEAR ANTIGEN (PCNA) AND
CC  REPLICATION FACTOR C (RFC) OR ACTIVATOR 1) FOR LEADING STRAND
CC  SYNTHESIS. ALSO INVOLVED IN COMPLETING OKAZAKI FRAGMENTS INITIATED
CC  BY THE DNA POLYMERASE ALPHA/PRIMASE COMPLEX (BY SIMILARITY).
CC  -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
CC  N PYROPHOSPHATE + DNA(N) -
CC  -1- SUBUNIT: HETERODIMER WITH SUBUNITS OF 125 KDA AND 50 KDA. THE 125
CC  KDA SUBUNIT CONTAINS THE POLYMERASE ACTIVE SITE AND MOST LIKELY
CC  THE ACTIVE SITE FOR THE 3'-5' EXONUCLEASE ACTIVITY (BY
CC  SIMILARITY).
CC  -1- SUBCELLULAR LOCATION: NUCLEAR.
CC  -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC  ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC  DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC  -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch)
-----
DR  EMBL; X88928; CAA61369.1; -.
DR  FLYBASE: FBgn0012066; DNAPol-delta.
DR  INTERPRO: IPR002064; -.
DR  PFAM: PF00136; DNA_POL_B; 1.
DR  PRINTS; PR00106; DNAPOLB.
DR  PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW  Transferase; DNA-directed DNA polymerase; DNA replication;
KW  DNA-binding; Hydrolase; Exonuclease; Zinc-finger; Nuclear protein.
FT  DOMAIN 4 19 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT  ZN_FING 997 1017 C4-TYPE (POTENTIAL).
FT  ZN_FING 1046 1064 C4-TYPE (POTENTIAL).
SQ  SEQUENCE 1092 AA; 124879 MW; 976B64D0BBB37DBD CRC64;

Query Match
Best Local Similarity 25.8%; Score 54; DB 1; Length 1092;
Matches 9; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

OY 13 YVGAKIGOVDAKOINGKNTAYG 34
    :||::|||  :|||  ||||
Db  692 FTGAOVGKIPCLIEISGSVYATG 713

RESULT 10
MOO_MYCTU
ID  MOO_MYCTU STANDARD: PRT: 493 AA.
AC  O05807;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-DEC-1998 (Rel. 37, Last sequence update)
DT  30-MAY-2000 (Rel. 39, Last annotation update)
DE  PROBABLE MALATE:QUINONE OXIDOREDUCTASE (EC 1.1.99.16) (MALATE
DE  DEHYDROGENASE [ACCEPTOR]) (MOO).
GN  MOO OR RV2852C OR MTCY24A1.05.

```



```

CC      SUBTILASE FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; U50194; AAA93458.1; -.
DR      HSSP; P00782; 25BT.
DR      MEROPS; S08_090; -.
DR      INTERPRO: IPR000209; -.
DR      PfAM; PF00082; Peptidase_S8; 3.
DR      PRINTS; PR00723; SUBTILASIN.
DR      PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
DR      PROSITE; PS00137; SUBTILASE_HIS; 1.
DR      PROSITE; PS00138; SUBTILASE_SER; 1.
DR      PROSITE; PS00290; IG_MIC; UNKNOWN_1.
KW      Hydrolyase.
FT      ACT_SITE      44      CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT_SITE      264     CHARGE RELAY SYSTEM (BY SIMILARITY).
FT      ACT_SITE      449     CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ      SEQUENCE      1249 AA; 138292 MW; F4A164028AA2B CRC64;

```

QY 10 AQPYVGAKIGQVDAKQINGKNTAYGIYAGY 39  
|||:|:|:|:  
Db 1138 AQPHDGAAGDAEAKEEHGGESTLESLEY 1167

QY 4 YGNSADAQPPYGA KIGVDAKQINGKN-TAYGIYAGYN 40  
|| | : || : : : || : : || : :  
Db 287 YGFPAAGAPPLGAL--HLDLRFNGKSWLVEGPGWGS 322

TPP2_RAT	STANDARD;	PRT;	1249(AA.
ID	TPP2_RAT		

DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE TRIPEPTIDYL-PEPTIDASE II (EC 3.4.14.10) (TPP-II) (TRIPEPTIDYL  
DE AMINOPEPTIDASE) (CHOLECYSTOKININ-INACTIVATING PEPTIDASE).  
CN TPP2.

05 Rattus norvegicus (Rat)

CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus

RN	[1]
RP	SEQUENCE FROM N.A.

RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;  
RX MEDLINE=96180116; PubMed=8602240;

RA Rose C., Vargas F., Facchinetti P., Bourgeat P., Bamba R.B.,  
RA Bishop P.B., Chan S.M., Moore A.N., Ganellin C.R., Schwartz J.C.;

RT "Characterization and inhibition of a cholecystokinin-inactivating  
RT serine peptidase.";

RL Nature 380:403-409(1996).  
CC -I- CATALYTIC ACTIVITY: RELEASE OF AN N-TERMINAL TRIPEPTIDE FROM A

CC POLYPEPTIDE.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).

CC

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8; ALSO KNOWN AS THE

```

CC SUBTILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U50194; AAA93458.1; -.
CC DR HSSP; P00782; 25BT.
CC DR MEROPS; S08.090; -.
CC DR INTERPRO; IPR000209; -.
CC DR PFAM; PF00082; Peptidase_S8; 3.
CC DR PRINTS; PR00723; SUBTILASTM.
CC DR PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG.
CC DR PROSITE; PS00137; SUBTILASE_HIS; 1.
CC DR PROSITE; PS00138; SUBTILASE_SER; 1.
CC DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
CC KW Hydrolyase.
CC FT ACT_SITE 44 44 CHANGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 264 264 CHANGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 449 449 CHANGE RELAY SYSTEM (BY SIMILARITY).
CC SQ SEQUENCE 1249 AA; 138292 MW; F4A41664028AA2B CRC64;
CC -----
QY 10 AQPYGAKIGQVDAKQINGKNTATITIVAGY 39
Db 1138 AQPHGGAAGDAEAKKEEGESTLEISERY 1167
-----
RESULT 12
UREG_HELPJ
ID UREG_HELPJ STANDARD; PRT; 199 AA.
AC Q9ZM27;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE UREASE ACCESSORY PROTEIN UREG.
GN UREG OR HpU068.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RA [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99120557; PubMed-9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Carnso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -!- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.
CC -!- SIMILARITY: BELONGS TO THE UREG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001445; AAD05647.1; -.
CC DR INTERPRO; IPR002894; -.
CC PFAM; PF01495; HypB_ureg; 1.

```

KW Nickel: ATP-binding. 15 ATP (POTENTIAL).  
FT NP\_BIND 8  
SQ SEQUENCE 199 AA; 21941 MW; 6959FFEB6A3D3C1 CRC64;  
Query Match 24.4%; Score 51; DB 1; Length 199;  
Best Local Similarity 42.3%; Pred. No. 10;  
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;  
OY 6 NSADAOPYVGA--KIGOVDAKOINCK 29  
DB 145 NKIDLAPYVGADLKVMERSKMRGE 170  
RESULT 13  
UREG\_HELPY STANDARD; PRT: 199 AA.  
AC 009066;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE UREASE ACCESSORY PROTEIN UREG.  
GN UREG OR HP0068.  
OS Helicobacter pylori (Campylobacter pylori).  
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
OC Helicobacter.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-85P.  
RX MEDLINE-92210488; PubMed-1313413;  
RA Cussac V., Ferrero R.L., Labigne A.;  
RT "Expression of Helicobacter pylori urease genes in Escherichia coli  
grown under nitrogen-limiting conditions.";  
RL J. Bacteriol. 174:2466-2473(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-26695 / ATCC 700392;  
RX MEDLINE-97394467; PubMed-9252185;  
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,  
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,  
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,  
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,  
RA McInerney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,  
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,  
RA Cotton M.D., Feldman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,  
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,  
RA Venter J.C.;  
RT "The complete genome sequence of the gastric pathogen Helicobacter  
pylori.";  
RL Nature 388:539-547(1997).  
CC -1- FUNCTION: PROBABLY FACILITATING NICKEL INCORPORATION.  
CC -1- SIMILARITY: BELONGS TO THE UREG FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
DR EMBL; M84338; AAA2025.1; -;  
DR EMBL; A24198; CAA01726.1; -;  
DR EMBL; AE000528; AAD07131.1; -;  
DR PIR; D41834; D41834.  
DR TIGR; HP0068; -;  
DR INTERPRO; IPR002894; -;  
DR PFAM; PF01495; Hypr\_ureg; 1.  
KW Nickel: ATP-binding.  
FT NP\_BIND 8  
FT CONFLICT 102 102 D -> S (IN REF. 1).  
FT CONFLICT 167 176 MRGKPFET -> IAAASPLFLP (IN REF. 1).  
SQ SEQUENCE 199 AA; 21955 MW; A23E448EB6A208B1 CRC64;

Query Match 24.4%; Score 51; DB 1; Length 199;  
Best Local Similarity 42.3%; Pred. No. 10;  
Matches 11; Conservative 6; Mismatches 7; Indels 2; Gaps 1;  
OY 6 NSADAOPYVGA--KIGOVDAKOINCK 29  
DB 145 NKIDLAPYVGADLKVMERSKMRGE 170  
RESULT 14  
YHIQ\_MEIGO STANDARD; PRT: 250 AA.  
ID YHIQ\_MEIGO  
AC P72077;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE HYPOTHETICAL 27.3 KDA PROTEIN.  
OS Neisseria gonorrhoeae.  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-M511;  
RA Carrick C.S., Pyfe J.A.M., Davies J.K.;  
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: STRONG, TO E.COLI YHIQ AND H.INFLUENZAE H10949.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC  
DR EMBL; U65994; AAC82508.1; -;  
DR KW Hypothetical protein.  
SQ SEQUENCE 250 AA; 27316 MW; FC749D6B91985763 CRC64;  
Query Match 24.2%; Score 50.5; DB 1; Length 250;  
Best Local Similarity 36.7%; Pred. No. 15;  
Matches 11; Conservative 6; Mismatches 12; Indels 1; Gaps 1;  
OY 11 OYVGAKI-GOVDAKQINKNTAVGIVAGY 39  
DB 216 RPRGEHLAQAPAYQYTGKSTREVDYLPY 245  
RESULT 15  
ID OPA\_HAEIN STANDARD; PRT: 121 AA.  
AC P45088;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE OPA PROTEIN.  
GN OPA OR H1174.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE-95350630; PubMed-7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RT Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT influenzae Rd.";  
RL Science 269:496-512(1995).  
RN [2]  
RP SEQUENCE OF 39-121 FROM N.A.  
RC STRAIN-RD / KW20;  
RX MEDLINE-96134971; PubMed-8550458;  
RA Preston A., Maskell D., Johnson A., Moxon E.R.;  
RT "Altered lipopolysaccharide characteristic of the 169 phenotype in  
RT Haemophilus influenzae results from mutations in a novel gene, isn.";  
RL J. Bacteriol. 178:396-402(1996).  
CC -1 - SIMILARITY: SOME TO N.GONORHOAE OPACITY PROTEIN OPA66.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U32797; AAC22827.1; -  
DR EMBL: U17295; AAA95981.1; -  
DR TIGR: H1174; -  
SQ SEQUENCE 121 AA; 13153 MW; 6647D8B471F42223 CRC64;

Query Match 23 9%; Score 50; DB 1; Length 121;  
Best Local Similarity 26.7%; Pred.No. 8.4;  
Matches 12; Conservative 6; Mismatches 17; Indels 10; Gaps 1;

OY 4 YGNSADAPYPYVGAKTIGOVDAKQINGK-----NTAVGIYAG 38  
: : |||||: | | | |  
Db 36 FDTNSKVQPYVGARVATNQFKYTNRAEQKFKSSSDIKIGYVAG 80

Search completed: March 6, 2001, 14:40:04  
Job time: 172 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:36:17 ; Search time 37.09 Seconds  
(without alignments)  
126.404 Million cell updates/sec

Title: US-09-164-714-1  
209  
Sequence: 1 A1SYGNSADAQPYGAKICGVDAKQJNGKNTAYGTYAGYN 40

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues  
Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: SP\_TREMBL\_15:\*  
2: sp\_archaea:\*  
3: sp\_bacteria:\*  
4: sp\_fungi:\*  
5: sp\_human:\*  
6: sp\_invertebrate:\*  
7: sp\_mhc:\*  
8: sp\_mammal:\*  
9: sp\_organelle:\*  
10: sp\_phage:\*  
11: sp\_plant:\*  
12: sp\_rodent:\*  
13: sp\_virus:\*  
14: sp\_vertebrate:\*  
14: sp\_unclassified:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63.5	30.4	407	2	P96774 haemophilus
2	59.5	28.5	380	5	O9XTP9 caenorhabditis
3	59	28.2	160	2	O9K575 vibrio chol
4	56	26.8	346	2	O51841 actinobacill
5	56	26.8	346	2	O9S539 actinobacill
6	55.5	26.6	767	3	O12357 saccharomyc
7	55.5	26.6	1155	3	O07163 saccharomyc
8	55.5	26.6	1196	3	O13527 saccharomyc
9	55.5	26.6	1285	3	O03934 saccharomyc
10	55.5	26.6	1328	3	O99231 saccharomyc
11	55.5	26.6	1328	3	O03855 saccharomyc
12	55.5	26.6	1346	3	O04345 saccharomyc
13	55.5	26.6	1346	3	O05679 saccharomyc
14	55.5	26.6	1346	3	O05369 saccharomyc
15	55.5	26.6	1347	3	O03494 saccharomyc
16	55.5	26.6	1749	3	O99337 saccharomyc
17	55.5	26.6	1755	3	O12088 saccharomyc
18	55.5	26.6	1755	3	O12112 saccharomyc
19	55.5	26.6	1755	3	O12141 saccharomyc

20	55.5	26.6	1755	3	O12269 saccharomyc
21	55.5	26.6	1755	3	O12273 saccharomyc
22	55.5	26.6	1755	3	O12316 saccharomyc
23	55.5	26.6	1755	3	O12414 saccharomyc
24	55.5	26.6	1755	3	O92393 saccharomyc
25	55.5	26.6	1755	3	O12490 saccharomyc
26	55.5	26.6	1755	3	O03612 saccharomyc
27	55.5	26.6	1755	3	O03619 saccharomyc
28	55.5	26.6	1756	3	O12193 saccharomyc
29	55.5	26.6	1770	3	O12113 saccharomyc
30	55.5	26.6	1770	3	O12501 saccharomyc
31	55.5	26.6	1770	3	O12503 saccharomyc
32	55.5	26.6	1770	3	O07791 saccharomyc
33	55.5	26.6	1770	3	O12472 saccharomyc
34	55.5	26.6	1770	3	O12491 saccharomyc
35	55.5	26.6	1771	3	P87006 saccharomyc
36	55.5	26.6	1793	3	O13535 saccharomyc
37	55.5	26.6	1810	3	O12022 saccharomyc
38	54.5	26.1	568	10	O9LXF5 haemophilus
39	54	25.8	403	2	P96773 streptomyces
40	54	25.8	612	2	O9KZ09 drosophila
41	54	25.8	1092	5	O9YUW8 drosophila
42	53.5	25.6	330	1	O27892 methanobact
43	53.5	25.6	1345	10	O64496 arabidopsis
44	53	25.4	259	10	O92T75 pinus taeda
45	53	25.4	748	5	O9V617 drosophila

## ALIGNMENTS

```

RESULT 1
ID P96774 PRELIMINARY: PRT: 407 AA.
AC P96774:
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE OMPA2.
GN OMPA2.
OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=730;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5000;
RX MEDLINE=97197543; PubMed=9045839;
RA Klesney-Tait J., Hiltke T.J., MacIver I., Spinoia S.M., Radolf J.D.,
RA Hansen E.J.;
RT "The major outer membrane protein of Haemophilus ducreyi consists of
RT two OmpA homologs.";
RL J. Bacteriol. 179:1764-1773(1997).
DR EMBL; U60646; AAB49274.1; -.
DR HSSP; P02934; 1BXW.
DR INTERPRO: IPR000498; -.
DR INTERPRO: IPR001145; -.
DR PFAM: PF00691; OmpA; 1.
DR PFAM: PF01389; OmpA_membrane; 1.
DR PRINTS: PRO1021; OMPADOMAIN.
DR PRODOM: PDD00930; ?; 1.
SO SEQUENCE 407 AA; 44702 MW; 0EBB62F3447F8C82 CRC64;

```

Query Match 30.4%; Score 63.5; DB 2; Length 407;  
Best Local Similarity 34.4%; Pred. No. 2;  
Matches 21; Conservative 3; Mismatches 8; Indels 29; Gaps 3;

```

QY 8 ADAPQ-----YVGAKIG-----QVDAKQJNGKNTA-----YGIYNG 38
DB 17 A1AIPADTFYVGAKAGWASFHRIQINQFDNRYKNRKHIVMEGKLNIRDSVTYGVFG 76
QY 39 Y 39

```

Db 77 Y 77

RESULT 2

09XTP9 PRELIMINARY: PRT: 380 AA.

AC 09XTP9; 01-NOV-1999 (TRENBLrel. 12, Created)

DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)

DE 01-JUN-2000 (TRENBLrel. 14, Last annotation update)

DE ZK1321.4 PROTEIN.

GN ZK1321.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Gardner A.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Willson R., Alnscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thelery-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.";

RL Nature 368:32-38(1994).

DR EMBL; Z48717; CA88612.1; -;

DR EMBL; Z48584; CA88612.1; JOINED.

DR EMBL; Z48584; CA88612.1; -;

DR EMBL; Z48717; CA88612.1; JOINED.

SO SEQUENCE 380 AA; 40259 MW; DC749476A488BF3 CRC64;

Query Match 28.5%; Score 59.5; DB 5; Length 380;

Best Local Similarity 38.6%; Pred. No. 6.3;

Matches 17; Conservative 2; Mismatches 16; Indels 9; Gaps 2;

OY 3 SYGNSADAPYVGAKIGVDKQIN-----GKNYAGIYAG 38

DB 224 SYNNGATSGPY-GAGSGGTPLNQMTFINTSPAPGANGAYGACG 266

RESULT 3

09KS75 PRELIMINARY: PRT: 160 AA.

AC 09KS75; 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE HYPOTHETICAL PROTEIN VC1384.

GN VC1384.

OS Vibrrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI\_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=EL TOR N16961 / SEROTYPE O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tellein H., Richardson D.,

RA Ermolaeva M.D., Yamathavan J., Bass S., Qin H., Dragol I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*

RT cholerae.";

RL Nature 406:477-483(2000).

DR EMBL; AE004217; AAF94542.1; -;

DR TIGR; VC1384; -;

KW Hypothetical protein.

SO SEQUENCE 160 AA; 16985 MW; 7E47CED047CE409D CRC64;

Query Match 28.2%; Score 59; DB 2; Length 160;

Best Local Similarity 45.2%; Pred. No. 2.7;

Matches 14; Conservative 6; Mismatches 9; Indels 2; Gaps 1;

OY 8 ADAOPYVGAKIGVDKQINKNYAGIYAG 38

DB 19 ADSWIYGASVGSQSDYEGKHG--TAYSVHAG 47

RESULT 4

051841 PRELIMINARY: PRT: 346 AA.

AC 051841; 01-JUN-1998 (TRENBLrel. 06, Created)

DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)

DT 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

DE OUTER MEMBRANE PROTEIN 34 PRECURSOR.

GN OMP34.

OS Actinobacillus actinomycetemcomitans

OS (Haemophilus actinomycetemcomitans).

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Actinobacillus.

OX NCBI\_TaxID=714;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=NCTC 9710;

RX MEDLINE=98084499; PubMed=9423883;

RA White P.A., Nair S.P., Kim M.J., Wilson M., Henderson B.;

RT "Molecular characterization of an outer membrane protein of

RT Actinobacillus actinomycetemcomitans belonging to the Ompa family.";

RL Infect. Immun. 66:369-372(1998).

DR EMBL; AF005079; AAC00068.1; -;

DR HSSP; P02934; 10TP.

DR INTERPRO; IPR000498; -;

DR INTERPRO; IPR011145; -;

DR PFAM; PF00691; OmpA; 1.

DR PFAM; PF01389; OmpA-membrane; 1.

DR PRINTS; PRO1021; OMPADOMAIN.

DR PROSITE; PS01068; OMPA; 1.

DR PRODOM; PD00930; -; 1.

KW Signal.

FT SIGNAL 1 21 POTENTIAL.

SO SEQUENCE 346 AA; 36905 MW; 68FFA4BC8B0F1819 CRC64;

Query Match 26.8%; Score 56; DB 2; Length 346;

Best Local Similarity 35.4%; Pred. No. 16;

Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;

OY 8 ADAOPYVGAKIGVDKQIN-----KQ-----INCKNTYAGIYAG 39

DB 19 AQAAPQANTFYAGAKAGWASSHHGLNDFKQKGVSTINRSEAYGVGG 66

RESULT 5

09S5J9 PRELIMINARY: PRT: 346 AA.

AC 09S5J9;

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE OMP29.
GN OMP34.
OS Actinobacillus actinomycetemcomitans
OS (Haemophilus actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC43718;
RA Komatsuzawa H., Kawai T., Wilson M.E., Taubman M.A., Sugai M.,
RA Suganaka H.;
RT "Cloning of the gene encoding the Actinobacillus actinomycetemcomitans
RT serotype beta OmpA-like outer membrane protein.";
RL Infect. Immun. 67:942-945(1999).
DR EMBL; AB015936; BAA75215.1; -.
DR INTERPRO; IPR000498; -.
DR INTERPRO; IPR001035; -.
DR INTERPRO; IPR001145; -.
DR INTERPRO; IPR002368; -.
DR PFAM; PF00691; OmpA; 1.
DR PFAM; PF01389; OmpA_membrane; 1.
DR PRINTS; PR01021; OMPADOMAIN.
DR PRINTS; PR01022; OUTERMURANA.
DR PRINTS; PR01023; NARLGMOY.
DR PROSITE; PS01068; OMPA; 1.
SQ SEQUENCE 346 AA; 36933 MW; 68E9D5DB8B0F1819 CRC64;
```

```
Query Match 26.8%; Score 56; DB 2; Length 346;
Best Local Similarity 35.4%; Pred. No. 16;
Matches 17; Conservative 4; Mismatches 11; Indels 16; Gaps 3;
```

```
QY 8 ADAQP-----YVGAKIGQVDA-----RQ-----INGKNTAYGIYAGY 39
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 19 AQAAPQANTFYAGAKAGMASHHGLNQPFQKGVSIKRNSEAGVGEFGY 66
```

```
RESULT 6
ID 012357 PRELIMINARY; PRT; 767 AA.
AC 012357;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE FRAME-SHIFT IN TYB PROBABLY NOT FUNCTIONAL.
GN TY1B.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rieger M., Mueller-Auer S., Brueckner M., Schaefer M.;
RN Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA MIPS;
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; Z72824; CAA97037.1; -.
DR EMBL; Z72823; CAA97029.1; -.
DR INTERPRO; IPR000194; -.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN.1.
SQ SEQUENCE 767 AA; 87951 MW; 5F69264A58605BA5 CRC64;
```

```
Query Match 26.8%; Score 55.5; DB 3; Length 767;
Best Local Similarity 43.8%; Pred. No. 49;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

```
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 624 SYGN----QPYKSQLGNIIFLNGKVIQCKST 651
```

```
RESULT 7
ID 007163 PRELIMINARY; PRT; 1155 AA.
AC 007163;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE (STRAIN JB84A CONTAINING PLASMID PNN162).
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JB84A;
RX MEDLINE=88246410; PubMed=2837641;
RA Boeke J.D., Eichinger D., Castillon D., Fink G.R.;
RT "The Saccharomyces cerevisiae genome contains functional and
RT nonfunctional copies of transposon Ty1.";
RL Mol. Cell. Biol. 8:1432-1442(1988).
DR EMBL; M18706; AAA66938.1; -.
DR INTERPRO; IPR001584; -.
DR PFAM; PF00665; rve; 1.
SQ SEQUENCE 1155 AA; 131935 MW; 2076198CD623484C CRC64;
```

```
Query Match 26.6%; Score 55.5; DB 3; Length 1155;
Best Local Similarity 43.8%; Pred. No. 80;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;
```

```
QY 3 SYGNSADAQPYVGAKIGQV---DAKQINGKNT 31
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1012 SYGN----QPYKSQLGNIIFLNGKVIQCKST 1039
```

```
RESULT 8
ID 013527 PRELIMINARY; PRT; 1196 AA.
AC 013527;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)
DE YAR009CP.
GN YAR009C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=85182712; PubMed=3886659;
RA Drabkin H.J., Rajbandary U.L.;
RT "Attempted expression of a human Initiator tRNA gene in Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 260:5596-5602(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=86028187; PubMed=2996783;
RA Hieter P., Pridmore D., Hegemann J.H., Thomas M., Davis R.W.,
RA Philippsen P.;
RT "Functional selection and analysis of yeast centromeric DNA.";
RL Cell 42:913-921(1985).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
```







```
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; IVE; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153931 MW; FF7F64800F9CAFE1 CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1346;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKQINGKNT 31
Db 1203 SYGN----QPYKSOIGNIFLNGKVIIGKST 1230

RESULT 14
005369 PRELIMINARY: PRT: 1346 AA.
AC 005369;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)
DE TRANSPOSON TY1-17 154.0 KDA HYPOTHETICAL PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]

SEQUENCE FROM N.A.
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae";
RL Nat. Genet. 10:261-268(1995).
DR EMBL: D50617; BAA09237.1; -.
DR INTERPRO: IPR00194; -.
DR INTERPRO: IPR001584; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; IVE; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1346 AA; 153890 MW; 2A71327338367C21 CRC64;

Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1346;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKQINGKNT 31
Db 1203 SYGN----QPYKSOIGNIFLNGKVIIGKST 1230

RESULT 15
003494 PRELIMINARY: PRT: 1347 AA.
AC 003494;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-OCT-2000 (TRENBLREL. 15, Last annotation update)
DE HYPOTHETICAL 153.9 KDA PROTEIN.
GN TYB.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA Oliver K., Shore L., Harris D.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-AB972;
RA Barrett B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL: Z68194; CAA92351.1; -.
DR INTERPRO: IPR00194; -.
DR INTERPRO: IPR001584; -.
DR INTERPRO: IPR001969; -.
DR PFAM: PF00665; IVE; 1.
DR PROSITE: PS00141; ASP_PROTEASE; 1.
DR PROSITE: PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR Hypothetical protein; Hydrolyase; Aspartyl protease.
SQ SEQUENCE 1347 AA; 153944 MW; 794BCFEB70B06FBB CRC64;
```

```
Query Match
Best Local Similarity 43.8%; Score 55.5; DB 3; Length 1347;
Matches 14; Conservative 5; Mismatches 6; Indels 7; Gaps 2;

QY 3 SYGNSADAOPYGAKIGOV---DAKQINGKNT 31
Db 1204 SYGN----QPYKSOIGNIFLNGKVIIGKST 1231
```

Search completed: March 6, 2001, 14:38:26  
Job time: 129 sec



22	39.4	7.3	1653	2	BOROSPAC	L23137	Borrellia	bu
23	39.4	7.3	1653	2	BOROSPAC	L23138	Borrellia	bu
24	39.4	7.3	1653	2	BOROSPAC	L23139	Borrellia	bu
25	39.4	7.3	1653	2	BOROSPAC	L23141	Borrellia	bu
26	39.4	7.3	1653	2	BOROSPAC	L23142	Borrellia	bu
27	39.4	7.3	1653	2	BOROSPAC	L23143	Borrellia	bu
28	39.4	7.3	1653	2	BOROSPAC	L23144	Borrellia	bu
29	39.4	7.3	1915	2	BHOSPAB	X14407	Borrellia	bu
30	39.4	7.3	1915	81	A0A009	A04009	B. burgdorferi	
31	39.4	7.3	1916	2	BOROSPAC	L19701	Borrellia	bu
32	39.4	7.3	1959	81	AR016731	AR016731	Sequence	
33	39.4	7.3	5952	81	AR064321	AR064321	Sequence	
34	39.4	7.3	53561	1	AE000790	AE000790	Borrellia	
35	39.2	7.2	195697	38	AC010707	AC010707	Drosophila	
36	39	7.2	136410	50	AC023308	AC023308	Homo sapi	
37	39	7.2	152355	9	AC005668	AC005668	Homo sapi	
38	38.8	7.2	198570	77	HS343C1	AL008720	Human DNa	
39	38.6	7.1	86064	6	AB011395	AB011395	Arabidops	
40	38.6	7.1	192518	54	AC027745	AC027745	Homo sapi	
41	38.6	7.1	197715	52	AC025493	AC025493	Homo sapi	
42	38.2	7.0	566	2	RCU45244	U45244	Rickettsia	
43	38.2	7.0	566	2	RCU46918	U46918	Rickettsia	
44	37.8	7.0	873	1	AF077663	AF077663	Borrellia	
45	37.8	7.0	891	2	BOROSPAC	L31399	Borrellia	bu

## ALIGNMENTS

	/gene="TAP1"	/note="Stop codon TAA-Q"	
misc_feature	1018..1020	/gene="TAP1"	
	/note="stop codon TGA"		
polyA_signal	1139..1144	/note="addition"	
BASE COUNT	414 a 122 c 317 g 311 t		
ORIGIN			
Query Match	8.3%	Score 45;	DB 56; Length 1164;
Best Local Similarity	50.7%;	Pred. No. 0.27;	
Matches 108; Conservative	0;	Mismatches 105;	Indels 0; Gaps 0;
OY	83 attctgctgatgctcaaccctatglttggtgccaaaattgyltcaagtagacgcgaagcaca	142	
db	653 ATGTGATGATGCCTCAAGACCGTGATGATGCCGGAAGATGGTGATGATTACGAAGATGCCG	712	
OY	143 tcaacggtaagaacacccgctatgatttatatgacggtataactttaccacaattttg	202	
Db	713 AAGATGGTGAATGTCCTAAGAATGGTGATGATGCCGGAAGATGGTGATGATTACAAGATGG	772	
OY	203 gcgtlaaaccccgaatttgttltcaagcgccaagaattaatlccaagcgtlaagtccgt	262	

RESULT	1
LOCUS	TTTAP1A 1164 bp mRNA INV 02-SEP-1994
DEFINITION	T.tetrahymena TAP1 gene polyA signal with multiple stop codons.
ACCESSION	X76126
VERSION	X76126.1 GI:426480
KEYWORDS	tapi gene.
SOURCE	Tetrahymena thermophila.
ORGANISM	Macronuclear Tetrahymena thermophila Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida; Tetrahymenina; Tetrahymena. 1 (bases 1 to 1164) Pearlman, R.E. Direct Submission Submitted (05-NOV-1993) R.E. Pearlman, York University, Dept of Biology, Farguharson BLDG., 4700 Keele St., Downsview, Ontario M3J 1P3, CANADA 2 (bases 1 to 1164) Heinonen, T.Y. and Pearlman, R.E. A germ line-specific sequence element in an Intron in Tetrahymena. J. Biol. Chem. 269 (26), 17428-17433 (1994) 94292495
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
MEDLINE	
FEATURES	Location/Qualifiers 1..1164 /organism="tetrahymena thermophila" /macronuclear /db_xref="taxon:5911" /chromosome="1" /dev_stage="macronucleus" /clone_lib="plasmid" /note="rearranged" 103..1020 /gene="TAP1" 103..103 /gene="TAP1" /note="stop codon TAA-Q" 961..963 /gene="TAP1" /note="stop codon TAA-Q" 967..969 /gene="TAP1" /note="stop codon TAA-Q" 997..999
misc_feature	
misc_feature	
misc_feature	
misc_feature	

Db	833	ATACGAAGATGCCGAAGATGCTGATGATGCTG	865
RESULT	2		
TTTAP1IN			
LOCUS	TTTAP1IN	1575 bp	DNA
DEFINITION	T. thermophila TAP1 gene.		INV
ACCESSION	X76125		29-SEP-1998
VERSION	X76125.1	GI:426479	
KEYWORDS	tap1 gene.		
SOURCE	Tetrahymena thermophila.		
ORGANISM	Macronuclear Tetrahymena thermophila		
	Eukaryota; Alveolata; Ciliophora; Oligohymenophorea;		
	Hymenostomatida; Tetrahymenina; Tetrahymena.		
	1 (bases 1 to 1575)		
REFERENCE	Pearlman, R. E.		
AUTHORS	Direct Submission		
TITLE	Submitted (05-NOV-1993) R.E. Pearlman, York University, Dept of		
JOURNAL	Biology, Farquharson Bldg., 4700 Keele St., Downsview, Ontario M3J		
	1P3, CANADA		
	2 (bases 1 to 1575)		
REFERENCE	Heinonen, T. Y. and Pearlman, R. E.		
AUTHORS	A gem. line-specific sequence element in an Intron in Tetrahymena		
TITLE	thermophila		
	J Biol. Chem. 269 (26), 17428-17433 (1994)		
	94292495		
JOURNAL	Location/Qualifiers		
MEDLINE	1..1575		
FEATURES	/organism="Tetrahymena thermophila"		
source	/macronuclear		
	/db_xref="taxon:5911"		
	/chromosome="1"		
	/dev_stage="macronucleus"		
	/clone_1lb="plasmid"		
	/note="rearranged"		
mrna	join(<16..205,383..664,899..>1431)		
gene	/gene="TAP1"		
	16..1555		
	/gene="TAP1"		
	<16..205		
exon	/gene="TAP1"		
	/number=1		
CDS	join(16..205,383..664,899..1431)		
	/gene="TAP1"		





MEDLINE 95179795  
 ERRATUM: [[published erratum appears in Curr Genet 1995  
 Apr;27(5):491]]  
 REFERENCE 2 (bases 1 to 823)  
 AUTHORS Pearlman, R.E.  
 TITLE Direct Submission  
 JOURNAL Submitted (28-FEB-1994) R.E. Pearlman, York University, Dept of  
 Biology, Fairquharson Bldg., 4700 Keele St., Downsview, Ontario M3J  
 1P3, CANADA

FEATURES  
 source Location/Qualifiers  
 1. 823

/organism="Candida albicans"  
 /strain="WO-1"  
 /db\_xref="taxon:5476"  
 /chromosome="1"  
 /clone.lib="plasmid Yep352"  
 /clone="pKs+"  
 /map="Sfi I fragment S"  
 149..746  
 /gene="DUT1"  
 /gene="DUT1"  
 149..156  
 /misc\_feature  
 /note="potential SCB element"  
 149..156  
 /gene="DUT1"  
 /note="Min I site (potential MCB element)"  
 197..200  
 /TATA\_signal  
 /gene="DUT1"  
 267..746  
 /gene="DUT1"  
 /codon\_start=1  
 /transl\_table=12  
 /product="dUTPase"  
 /protein\_id="CAA54897.1"  
 /db\_xref="GI:457721"  
 /db\_xref="SWISS-PROT:P43058"  
 /translation="MTSEDSLKQKLESTOSLKYLNSPKKVPKTSALAGYDLY  
 SAEAVTPAHGGLVSTDISIVPIGTGVAPRSLAVKHISTGAGVIDADYRGVEV  
 KVLFNHNSKDEIKEDRIADQLVEOIVNADIKELISELDNTERGEGCGSTGKN"  
 BASE COUNT 290 a 138 c 146 g 249 t  
 ORIGIN

Query Match 8.0%; Score 43.6; DB 7; Length 823;  
 Best Local Similarity 46.2%; Pred. No. 0.62;  
 Matches 145; Conservative 0; Mismatches 169; Indels 0; Gaps 0;

67 gccatcagctatgcaattctgctgatacgaacctatgctgccaattgtcaaa 126  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 510 GCCGTGAACACATGTTTCACACTGGTGTGTCATTCGATGCTGATTTATCGTGTGA 569  
 127 gtacagcgcaagaacaatcagcgttaagacacgcgttatgctatgacgattat 186  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 570 GTAAAGTAGTCTTATTATTAACATAGTGAAGAAAGATTTCGAATTTAAGAGGTATAGA 629  
 187 tttagacaaatttggcgttagaaccgaattgtgtgtcagagcgccaagaattaat 246  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 630 ATTGCTCAATTTGATTTGGAACAATCGTCATGTCATCAAGAATTTCACTTGAA 689  
 247 gcaagcgtagtcctcgttaaaaggtgatgtgaagtccttttggctctatgacacatcgc 306  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 690 GAATTAGATTAATCTGACACTGTAAGAGGTGTTTGATCTACTGTTAAGAACTAGGAC 749  
 307 tacaacttcaatacccaattttagccaagcgcaaataggcatctgtaagactaaa 366  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 750 TACTACTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATCGTA 809  
 367 gtatagcttaccag 380  
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111  
 DB 810 TTCTCTATAAAGAG 823

RESULT 5

AC008125  
 LOCUS  
 DEFINITION  
 ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 COMMENT

AC008125 189286 bp DNA PRI 25-SEP-1999  
 Homo sapiens 12 BAC RPc11-25E2 (Roswell) Park Cancer Institute  
 Human BAC Library) complete sequence.

AC008125  
 AC008125.9 GI:5923640  
 HTG.

human.  
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 189286)  
 Muzny/D.M., Adams/C., Bailey/M., Barbara/J., Blankenburg/K.,  
 Bodola/B., Bouck/J., Bowler/S., Brooks/A., Bunay/C., Bunac/C.,  
 Burkett/C., Burrows/J., Carter/M., Chacko/J., Chen/Z., Cox/C.,  
 David/R., Delgado/O., Deshazo/D., Ding/Y., Domah-Rashid,N.,  
 Dugan-Rocha/S., Durbin/K.J., Fernandez/C., Ferriguto/G.,  
 Forcum-Tansey/J., Frantz/P., Ganes/R., Gorrell/J.H., Gorrell,L.L.,  
 Guenvera/M., Harris/K., Hernandez/J., Hodgson,A., Hogues/M.,  
 Hollaway/C., Hosak,H., Issar,A., Jackson,L.E., Jackson,L., Jia,Y.,  
 Jones,M., Kelly/S., Kneitz/S., Kondajewski,N., Kong,Y., Kovar,C.,  
 Lau,S., Leal,B., Lee,E., Li,Z., Lichteberg,O., Liu,J., Liu W.,  
 Logan,O., Lu,J., Lucier/R., Marondel,I., Martin,R., Martinez,C.,  
 McLeod,M.P., Mei/G., Merscher/S., Miller,A., Montgomery,K.T.,  
 Morgan,M., Morris/S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N.,  
 Nguyen,S., Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L.,  
 Pu,L.L., Quiles/M., Reiter,D., Rives/M., Samuel,S., Say,J.,  
 Scherer,S., Shah,E., Shen,H., Shim,C., Simon,M., Sparks,A.,  
 Stamps,A., Sugang,R., Tabot,P., Taylor,T., Vasquez,L., Vinson,R.,  
 Vo,Q., Wahbah,M., Watlington,S., Weinstein/G., Weinstein,I.R.,  
 Williamson,A., Worley,K., Wren,J., Wrensford,G., Xiang,A.M.,  
 Yang,R., Yu,W., Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.  
 Direct Submission  
 2 (bases 1 to 189286)  
 Worley,K.C.  
 Direct Submission  
 Submitted (24-JUL-1999) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Sep 25, 1999 this sequence version replaced gi:5651701.  
 INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email  
 gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the  
 entire insert of this clone. Overlapping regions of clones are only  
 sequenced and submitted once, so the sequence for the remainder of  
 the insert may be found in the record for the adjacent clones.  
 Overlapping clones are noted at the beginning and end of the  
 features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches  
 of a local database that includes entries from dbSTS, GDB, and  
 local mapping efforts.  
 Repeats are identified using RepeatMasker (A. Smit and P. Green,  
 unpublshed.) for Human and Mouse sequences.  
 Genes and Region of sequence similarity are identified by BLAST  
 (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
 EST and cDNA sequences. Genes demonstrate at least two exons  
 flanked by consensus splice sites that maintained sequence  
 continuity across the splice junctions. Sequences that are not  
 identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
 standard of double strand coverage with a minimum of 2 clones and 2  
 reads with no ambiguities or 2 chemistries with a minimum of 2

clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: [http://gc.bcm.tmc.edu:8086/quality\\_info/genbank\\_annotation.html](http://gc.bcm.tmc.edu:8086/quality_info/genbank_annotation.html).

## QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length: 189286
Phrap values in estimate: 188412
Average error rate (BCM-Phrap estimate): 0.000392177
Fraction of Phrap values less than 40 : 0.0664395
Number of consensus changing edits: 22
Number of N's in consensus : 0
```

```
----- Consensus changing edits -----
Position Original+Context Edited+Context
5897 atcacacct(n)caattccct atcacacct(c)caagttccct
42337 aatgcttct(n)caagctcagt aatgcttct(c)caagctcagt
47267 ttltttaga(n)ggagctctgc ttltttaga(t)ggagctctgc
55095 lgtatgaag(c)n)gtlctcaatc lgtatgaag(c)gtlctcaatc
74098 cagcttctct(n)nnctccctt cagcttctct(g)ctgcccctt
74099 agcttctct(n)ntccctt agcttctct(g)ctgcccctt
74100 gcttctctm(n)ntccctt gcttctctm(t)gtcctctta
74101 ctctctctm(n)lcccccttac ctctctctg(c)lcccccttac
77151 lgaagaatg(n)anagataaac lgaagaatg(a)aaagataaac
77153 aaagaatgna(n)agataataaa aaagaatgna(a)agataataaa
95463 ttgagagagc(n)ccacaggtca ttgagagagc(a)ccacaggtca
103775 gtacacatag(a)atacatgtgc gtacacatag(c)atacatgtgc
104001 atgagtgtt(n)tagcttcac atgagtgtt(c)tagcttcac
151027 cccatgaac(n)caaatatcac cccatgaac(a)caaatatcac
151031 agaaacnac(n)aattactaaa agaaacnac(a)aattactaaa
151057 clcaagagga(n)aaagaaatc clcaagagga(a)aaagaaatc
151154 ccaaatgtt(n)aaagaaat ccaaatgtt(a)aaagaaat
151643 cccacagtaa(n)aatcacagag cccacagtaa(g)aatcacagag
173642 tatatatag(n)tatatatata tatatatag(a)tatatatata
173684 atatatata(n)ctatatata atatatata(a)ctatatata
179756 atgtatctt(n)tgagacacac atgtatctt(t)tgagacacac
186035 ggaagcccca(n)ccaacacatg ggaagcccca(g)ccaacacatg
```

## ----- Distribution of Quality &lt; 40 Bases -----

#	bases	5	10	15	20	25	30	35	40
10001					*	*	*	*	*
9001					*	*	*	*	*
8001					*	*	*	*	*
7001					*	*	*	*	*
6001					*	*	*	*	*
5001					*	*	*	*	*
4001				*	*	*	*	*	*
3001			*	*	*	*	*	*	*
2001		*	*	*	*	*	*	*	*
1001		*	*	*	*	*	*	*	*
01		*	*	*	*	*	*	*	*

Version: 1.01 gxiO.  
Location/Qualifiers  
1. .189286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

```
/chromosome="12"
/clone="RPC111-25E2"
58. .174
/rpt_family="MER103"
complement(309. .376)
/rpt_family="Alu"
1586. .1653
/rpt_family="MIR"
complement(1693. .1817)
/rpt_family="MIR"
1882. .2059
/rpt_family="L2"
complement(2681. .2813)
/rpt_family="L2"
complement(3681. .3898)
/rpt_family="MIR"
4057. .4368
/rpt_family="AluSp"
6005. .6025
/rpt_family="(A)n"
complement(6334. .6461)
/rpt_family="MIR"
6619. .7190
/rpt_family="L2"
complement(7191. .7492)
/rpt_family="AluX"
7493. .7706
/rpt_family="L2"
8149. .8411
/rpt_family="AluX"
8414. .8731
/rpt_family="AluX"
8746. .8846
/rpt_family="GA-rich"
8860. .9153
/rpt_family="AluJb"
complement(9162. .9447)
/rpt_family="MER96B"
9448. .9597
/rpt_family="MIR"
10481. .10661
/rpt_family="L2"
10667. .10695
/rpt_family="(TAA)n"
11027. .11113
/rpt_family="L1MD1"
complement(11118. .11522)
/rpt_family="MER57B"
11536. .11565
/rpt_family="AT-rich"
11819. .11949
/rpt_family="L2"
complement(12057. .12360)
/rpt_family="AluX"
12772. .12968
```

Query Match 8.0%; Score 43.6; DB 9; Length 189286;  
Best Local Similarity 47.7%; Pred. No. 0.68;  
Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```
QY 56 gtcctaagctccatgaagctatggaattctgcgtagtctaaccctatgtgtgcca 115
DB 26855 GTGATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 26914
QY 116 aatgtgtcaagtagaagccaagcaaatcaaggttaagaacccgctatglatlata 175
DB 26915 ATGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGATG 26974
QY 176 caggtataactlttgaccaaactlttgcgtagaaccgcaatlttgltcagacgcca 235
DB 26975 ATGTTATGTCAGTGCATGCTGATGATGATGATGATGATGATGATGATGATG 27034
QY 236 aagaattaatgcagcgtagtccctcttaaaagtgatgtgaagtccttggctctatg 295
```





ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 213140)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 1, clone RP11-217N8  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 213140)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearrellano,K., Dewar,K., Diaz,J.S.,  
Dodgess,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,  
McCarthy,M., McEwan,P., McGuck,A., McKernan,K., McPheters,R.,  
Meldrum,J., Meneses,L., Milhova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Sudramanian,A., Talamas,J.,  
Teefaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,C., Zainoun,J., Zimmer,A. and Zody,M.  
Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L7715  
Center clone name: 217.N.8  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 182556 bases at least Q40  
Consensus quality: 198779 bases at least Q30  
Consensus quality: 204626 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 207940; sum-of-ctnigs  
Quality coverage: 3.9 in Q20 bases; agarose-fp  
Quality coverage: 3.4 in Q20 bases; sum-of-ctnigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 53 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
\* 1 1214: contig of 1214 bp in length  
\* 1215 1314: gap of 100 bp  
\* 1315 2372: contig of 1058 bp in length  
\* 2373 2472: gap of 100 bp  
\* 2473 3493: contig of 1023 bp in length  
\* 3496 3593: gap of 100 bp  
\* 3596 4625: contig of 1030 bp in length  
\* 4626 4725: gap of 100 bp  
\* 4726 5839: contig of 1114 bp in length  
\* 5840 5939: gap of 100 bp

\* 5940 7014: contig of 1075 bp in length  
\* 7015 7114: gap of 100 bp  
\* 7115 8180: contig of 1066 bp in length  
\* 8181 8280: gap of 100 bp  
\* 8281 9604: contig of 1324 bp in length  
\* 9605 9704: gap of 100 bp  
\* 9705 11339: contig of 1635 bp in length  
\* 11340 11439: gap of 100 bp  
\* 11440 12708: contig of 1269 bp in length  
\* 12709 12808: gap of 100 bp  
\* 12809 14202: contig of 1304 bp in length  
\* 14203 14302: gap of 100 bp  
\* 14303 15487: contig of 1185 bp in length  
\* 15488 15587: gap of 100 bp  
\* 15588 16909: contig of 1322 bp in length  
\* 16910 17009: gap of 100 bp  
\* 17010 18251: contig of 1242 bp in length  
\* 18252 18351: gap of 100 bp  
\* 18352 19767: contig of 1416 bp in length  
\* 19768 19867: gap of 100 bp  
\* 19868 21172: contig of 1305 bp in length  
\* 21173 21272: gap of 100 bp  
\* 21273 23727: contig of 2455 bp in length  
\* 23728 23827: gap of 100 bp  
\* 23828 24931: contig of 1104 bp in length  
\* 24932 25031: gap of 100 bp  
\* 25032 26475: contig of 1444 bp in length  
\* 26476 26575: gap of 100 bp  
\* 26576 28259: contig of 1664 bp in length  
\* 28260 28359: gap of 100 bp  
\* 28360 31078: contig of 2719 bp in length  
\* 31079 31178: gap of 100 bp  
\* 31179 32716: contig of 1538 bp in length  
\* 32717 32816: gap of 100 bp  
\* 32817 33881: contig of 1065 bp in length  
\* 33882 33981: gap of 100 bp  
\* 33982 36218: contig of 2237 bp in length  
\* 36219 36318: gap of 100 bp  
\* 36319 37461: contig of 1143 bp in length  
\* 37462 37561: gap of 100 bp  
\* 37562 40250: contig of 2669 bp in length  
\* 40251 40350: gap of 100 bp  
\* 40351 41626: contig of 1276 bp in length  
\* 41627 41726: gap of 100 bp  
\* 41727 44548: contig of 2822 bp in length  
\* 44549 44648: gap of 100 bp  
\* 44649 46762: contig of 2114 bp in length  
\* 46763 46862: gap of 100 bp  
\* 46863 49156: contig of 2294 bp in length  
\* 49157 49256: gap of 100 bp  
\* 49257 51320: contig of 2064 bp in length  
\* 51321 51420: gap of 100 bp  
\* 51421 53478: contig of 2058 bp in length  
\* 53479 53578: gap of 100 bp  
\* 53579 56891: contig of 3313 bp in length  
\* 56892 56991: gap of 100 bp  
\* 56992 59163: contig of 2172 bp in length  
\* 59164 59263: gap of 100 bp  
\* 59264 62101: contig of 2838 bp in length  
\* 62102 62201: gap of 100 bp  
\* 62202 65836: contig of 3635 bp in length  
\* 65837 65936: gap of 100 bp  
\* 65937 69543: contig of 3607 bp in length  
\* 69544 69643: gap of 100 bp  
\* 69644 74765: contig of 5122 bp in length  
\* 74766 74865: gap of 100 bp  
\* 74866 78671: contig of 3806 bp in length  
\* 78672 78771: gap of 100 bp  
\* 78772 83032: contig of 4261 bp in length  
\* 83033 83132: gap of 100 bp  
\* 83133 90972: contig of 7840 bp in length  
\* 90973 91072: gap of 100 bp  
\* 91073 97926: contig of 6854 bp in length

```

* 97927 98026: gap of 100 bp
* 98027 105572: contig of 7546 bp in length
* 105573 105672: gap of 100 bp
* 105673 113386: contig of 7714 bp in length
* 113387 113486: gap of 100 bp
* 113487 121933: contig of 8447 bp in length
* 121934 122033: gap of 100 bp
* 122034 131116: contig of 9083 bp in length
* 131117 131216: gap of 100 bp
* 131217 139837: contig of 8621 bp in length
* 139838 139937: gap of 100 bp
* 139938 148533: contig of 8596 bp in length
* 148534 148633: gap of 100 bp
* 148634 157076: contig of 8443 bp in length
* 157077 157176: gap of 100 bp
* 157177 169834: contig of 12658 bp in length
* 169835 169934: gap of 100 bp
* 169935 182103: contig of 12169 bp in length
* 182104 182203: gap of 100 bp
* 182204 197520: contig of 15317 bp in length
* 197521 197621: gap of 100 bp
* 197621 213140: contig of 15520 bp in length.

```

## FEATURES

```

source
  1.213140
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="1"
  /map="1"
  /clone_lib="RP11-217N8"
  /clone_1ib="RP11-217N8"
  /note="assembly_fragment"
  1.1214
  /note="assembly_fragment"
  1315.2372
  /note="assembly_fragment"
  2473.3495
  /note="assembly_fragment"
  3596.4625
  /note="assembly_fragment"
  4726.5839
  /note="assembly_fragment"
  5940.7014
  /note="assembly_fragment"
  7115.8180
  /note="assembly_fragment"
  8281.9604
  /note="assembly_fragment"
  9705.11339
  /note="assembly_fragment"

```

Query Match 8.0%; Score 43.6; DB 54; Length 213140;

Best Local Similarity 47.7%; Pred. No. 0.68;

Matches 127; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

```

QY 56 gtctaacgtgcacatcagctatgcacatctctgctgctcaacctatggtgtcga 115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188108 GTGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 188049

QY 116 aaattgtcagaagacgcaagcaaaatcaacggttaagaacaccgcttatgtaattat 175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188048 ATGATGCTTATGCTGATGATGATGATGCTTATGCTGATGATGCTGATGATG 187989

QY 176 cagggtataaacttgcacaaatttggcgttagaaccggaatttgggttcagaagcga 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187988 ATGCTTATGCTGATGCTGATGCTGATGCTTATGCTGATGCTGATGCTGATG 187929

QY 236 aagaatttaatgcagcgctagtcctgtaaaagtgatggaagcttttggtgcttatg 295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187928 ATGGTATGACGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 187869

QY 296 gcaacatacgtataactatcatcat 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187868 GTATATGATGATTTATATATGATGATT 187843

```

## RESULT 8

AC026243 134741 bp DNA HNG 13-JUL-2000  
Homo sapiens clone RP11-25G3, LOW-PASS SEQUENCE SAMPLING.  
AC026243.2 GI:7341988  
HTG: HTGS\_PHASE0.  
human.

## REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 134741)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

## AUTHORS

Homo sapiens, clone RP11-25G3  
Unpublished  
2 (bases 1 to 134741)

## TITLE

Unpublished

## REFERENCE

Unpublished

## AUTHORS

Unpublished

## TITLE

Submitted (21-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Mar 30, 2000 this sequence version replaced g1:7272142.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: MIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
Project Information  
Center project name: L4567  
Center clone name: 25\_G\_3

NOTE: This record contains 150 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
796 895: contig of 795 bp in length  
796 895: gap of 100 bp  
896 1674: contig of 779 bp in length  
1675 1774: gap of 100 bp  
1775 2566: contig of 792 bp in length  
2567 2666: gap of 100 bp  
2667 3453: contig of 787 bp in length  
3454 3553: gap of 100 bp  
3554 4368: contig of 815 bp in length

\* 4469 4468: gap of 100 bp  
\* 4469 5268: contig of 800 bp in length  
\* 5269 5368: gap of 100 bp  
\* 5369 6147: contig of 779 bp in length  
\* 6148 6247: gap of 100 bp  
\* 6248 7036: contig of 789 bp in length  
\* 7037 7136: gap of 100 bp  
\* 7137 7922: contig of 786 bp in length  
\* 7923 8022: gap of 100 bp  
\* 8023 8819: contig of 797 bp in length  
\* 8820 8919: gap of 100 bp  
\* 8920 9722: contig of 803 bp in length  
\* 9723 9822: gap of 100 bp  
\* 9823 10604: contig of 782 bp in length  
\* 10605 10704: gap of 100 bp  
\* 10705 11466: contig of 762 bp in length  
\* 11467 11566: gap of 100 bp  
\* 11567 12363: contig of 797 bp in length  
\* 12364 12463: gap of 100 bp  
\* 12464 13250: contig of 787 bp in length  
\* 13251 13350: gap of 100 bp  
\* 13351 14130: contig of 780 bp in length  
\* 14131 14230: gap of 100 bp  
\* 14231 15030: contig of 800 bp in length  
\* 15031 15130: gap of 100 bp  
\* 15131 15928: contig of 798 bp in length  
\* 15929 16028: gap of 100 bp  
\* 16029 16809: contig of 781 bp in length  
\* 16810 16909: gap of 100 bp  
\* 16910 17701: contig of 792 bp in length  
\* 17702 17801: gap of 100 bp  
\* 17802 18586: contig of 785 bp in length  
\* 18587 18686: gap of 100 bp  
\* 18687 19502: contig of 816 bp in length  
\* 19503 19602: gap of 100 bp  
\* 19603 20406: contig of 804 bp in length  
\* 20407 20506: gap of 100 bp  
\* 20507 21275: contig of 769 bp in length  
\* 21276 21375: gap of 100 bp  
\* 21376 22168: contig of 793 bp in length  
\* 22169 22268: gap of 100 bp  
\* 22269 23073: contig of 805 bp in length  
\* 23074 23173: gap of 100 bp  
\* 23174 23974: contig of 801 bp in length  
\* 23975 24074: gap of 100 bp  
\* 24075 24839: contig of 765 bp in length  
\* 24840 24939: gap of 100 bp  
\* 24940 25705: contig of 766 bp in length  
\* 25706 25805: gap of 100 bp  
\* 25806 26600: contig of 795 bp in length  
\* 26601 26700: gap of 100 bp  
\* 26701 27488: contig of 788 bp in length  
\* 27489 27588: gap of 100 bp  
\* 27589 28373: contig of 785 bp in length  
\* 28374 28473: gap of 100 bp  
\* 28474 29278: contig of 805 bp in length  
\* 29279 29378: gap of 100 bp  
\* 29379 30187: contig of 809 bp in length  
\* 30188 30287: gap of 100 bp  
\* 30288 31062: contig of 775 bp in length  
\* 31063 31162: gap of 100 bp  
\* 31163 31949: contig of 787 bp in length  
\* 31950 32049: gap of 100 bp  
\* 32050 32841: contig of 792 bp in length  
\* 32842 32941: gap of 100 bp  
\* 32942 33725: contig of 784 bp in length  
\* 33726 33825: gap of 100 bp  
\* 33826 34633: contig of 808 bp in length  
\* 34634 34733: gap of 100 bp  
\* 34734 35515: contig of 782 bp in length  
\* 35516 35615: gap of 100 bp  
\* 35616 36404: contig of 789 bp in length  
\* 36405 36504: gap of 100 bp

\* 36505 37288: contig of 784 bp in length  
\* 37289 37388: gap of 100 bp  
\* 37389 38185: contig of 797 bp in length  
\* 38186 38285: gap of 100 bp  
\* 38286 39090: contig of 805 bp in length  
\* 39091 39190: gap of 100 bp  
\* 39191 39985: contig of 795 bp in length  
\* 39986 40085: gap of 100 bp  
\* 40086 40867: contig of 782 bp in length  
\* 40868 40967: gap of 100 bp  
\* 40968 41747: contig of 780 bp in length  
\* 41748 41847: gap of 100 bp  
\* 41848 42615: contig of 768 bp in length  
\* 42616 42715: gap of 100 bp  
\* 42716 43495: contig of 780 bp in length  
\* 43496 43595: gap of 100 bp  
\* 43596 44384: contig of 789 bp in length  
\* 44385 44484: gap of 100 bp  
\* 44485 45281: contig of 797 bp in length  
\* 45282 45381: gap of 100 bp  
\* 45382 46189: contig of 808 bp in length  
\* 46190 46289: gap of 100 bp  
\* 46290 47085: contig of 796 bp in length  
\* 47086 47185: gap of 100 bp  
\* 47186 47956: contig of 771 bp in length  
\* 47957 48056: gap of 100 bp  
\* 48057 48846: contig of 790 bp in length  
\* 48847 48946: gap of 100 bp  
\* 48947 49745: contig of 799 bp in length  
\* 49746 49845: gap of 100 bp  
\* 49846 50629: contig of 784 bp in length  
\* 50630 50729: gap of 100 bp  
\* 50730 51519: contig of 790 bp in length  
\* 51520 51619: gap of 100 bp  
\* 51620 52398: contig of 779 bp in length  
\* 52399 52498: gap of 100 bp  
\* 52499 53250: contig of 752 bp in length  
\* 53251 53350: gap of 100 bp  
\* 53351 54154: contig of 804 bp in length  
\* 54155 54254: gap of 100 bp  
\* 54255 55065: contig of 811 bp in length  
\* 55066 55165: gap of 100 bp  
\* 55166 55949: contig of 784 bp in length  
\* 55950 56049: gap of 100 bp  
\* 56050 56854: contig of 805 bp in length  
\* 56855 56954: gap of 100 bp  
\* 56955 57752: contig of 798 bp in length  
\* 57753 57852: gap of 100 bp  
\* 57853 58691: contig of 839 bp in length  
\* 58692 58791: gap of 100 bp  
\* 58792 59590: contig of 799 bp in length  
\* 59591 59690: gap of 100 bp  
\* 59691 60510: contig of 820 bp in length  
\* 60511 60610: gap of 100 bp  
\* 60611 61410: contig of 800 bp in length  
\* 61411 61510: gap of 100 bp

Query Match 7.8%; Score 42.4; DB 53; Length 134741;

Best Local Similarity 47.7%; Pred. No. 1.4;

Matches 124; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 56 grgctaacgcctgcacatcagctatgcaatctcgtcgtatgctcaaccctatgtgtgca 115  
DB 89095 gTGCATGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 89036  
QY 116 aaattggtcaagtaagcgcgaagcaatacaagtaagaacccgtttgtatattatg 175  
DB 89035 ATGATGCTTATGCTGCTGATGATGATGCTTATGCTGCTGATGCTGATGCTGATG 88976  
QY 176 caggtataactttagaccaaaatlilgycglagaaacccgaatttltgttcaagcaca 235  
DB 88975 ATGCTTATGCGACACTGACGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 88916





```

/product="G4p2"
/protein_id="AAA70169.1"
/db_xref="GI:710338"
/translation="MSNPEDLLGNVEDADVVLPKEIKVSNSTSSKKADVPSPADP
/KKRNRRPPSGNMGARIRDKTGRNRNRKSDYTDSTKSNRRRAVDHSHRTGKDT
/KKVVQMGDDKKEI/SAEKEADADAELAEADAEEDGKPKTALSLQDYLNOQAN
/NOENVPKREVELDAGRIETAEKAYVATVKNVSKQLTKRELEADAFVESNT
/RNFGDRNRNSNNFNRRGRGARGKRNNTANATNSANTVQRNRNIDVSNLPSLA"

```

```

BASE COUNT      415 a      262 c      250 g      350 t
ORIGIN

```

```

Query Match      7.5%; Score 40.8; DB 47; Length 1277;
Best Local Similarity 44.7%; Pred. No. 3.3;
Matches 159; Conservative 0; Mismatches 197; Indels 0; Gaps 0;

```

```

QY  91 gatgtcaaacctatgttggtgccaataatgtgcgaagtgaagccgaacaaatcaaggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  597 GAAGCTCAAGCCGATCTGCTGTAATTGCTGAAGACCTCAGAAAGCTGAAGAGCT 656
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 aagaacacccgttatgtatattatgcaggtataactctgacaaaatttgcgtagaa 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  657 GGAAGCCAAAGACCCCTCAATTGCTTTGCAAGACTACTTGAACCAACAGCTAACAC 716
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 cccgaatttggttcagaagcgaagaatlaatgcagcgtagagtcctgtaaaagt 270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  717 CAGTTCACAGAGTCCAGACGAAGAGAGCTGAATTGACCGCTGGAAGAAATTGAAC 776
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  271 gatgtgaagtctttgtgtctatgcacatcgcctataacttcatacccaatt 330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  777 GCTGAAGAGAGAGACTTACGTTCCAGCAACCAAGCTGACAGCTCAATTAAGCAATTG 836
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  331 tatgcgaagggcaaatatgagcattgctgaagtaagtagatgttaccagccgtaagca 390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  837 AAGACCAAGAGAGACTTGCAATTTCATGACACTTTTGTTAATCACTAGAAAAGAAC 896
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  391 actacatactcaaaaagcgaacaaacagcctagcagcggtgtgtgtgtg 446
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  897 TTGCGTGACAGAAACAACACAGCAGAAACACTTCAACACGCTGCTGTGAG 952
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 13
AC074318/c      51788 bp      DNA      12-OCT-2000
LOCUS           Staphylococcus aureus clone sabac-108, WORKING DRAFT SEQUENCE, 1
DEFINITION
ORDERED_PIECES:
AC074318      51788 bp      DNA      12-OCT-2000
AC074318      GI:10799432
VERSION
KEYWORDS       HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE
ORGANISM       Staphylococcus aureus
                Staphylococcus aureus
                Bacillus/Clostridium group;
                Bacillus/Staphylococcus group; Staphylococcus.
                1 (bases 1 to 51788)
                Tian, R., Lin, S., Jia, H., Qian, Y., Iondola, J. and Roe, B.A.
                Staphylococcus aureus BAC Clone sabac-108
                Unpublished
                2 (bases 1 to 51788)
                Tian, R., Lin, S., Jia, H., Qian, Y., Iondola, J. and Roe, B.A.
                Direct Submission
                Submitted (26-JUL-2000) Department Of Chemistry And Biochemistry,
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA

```

On Oct 12, 2000 this sequence version replaced gi:10765076.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 1 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* provided by the submittor.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

```

*
FEATURES
  1 51788: Contig of 51788 bp in length.
    Location/Qualifiers
      source
        /organism="Staphylococcus aureus"
        /db_xref="taxon:1280"
        /clone="sabac-108"
        /clone_lib="OU staph library"
        /clone_id="9849 c 7959 g 17522 t"
BASE COUNT      16458 a      9849 c      7959 g      17522 t
ORIGIN

```

```

Query Match      7.4%; Score 40.4; DB 59; Length 51788;
Best Local Similarity 51.7%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

```

QY  91 gatgtcaaacctatgttggtgccaataatgtgcgaagtgaagccgaacaaatcaaggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  3119 GATGCTACGTATTCTATGTAAGACATGGGTACCTTTATACCAACAGCTGAAATTG 3060
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  151 aagaacacccgttatgtatattatgcaggtataactctgacaaaatttgcgtagaa 210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  3059 GAAACCAAGCGCGACCTTTTATGAAGGTATGAAGTGAAGTGAAGCAAGCA 3000
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY  211 cccgaatttggttcagaagcgaagaatlaatgcagcgtagagtcctgtaaaag 268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  2999 AAACCTTGATGTGATGAGTGTAAATGATGACACAGCTGTGAAACCGAACAG 2942
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 14
AC025591      58431 bp      DNA      BCT      01-JUL-2000
LOCUS           Staphylococcus aureus clone sabac-134, complete sequence.
DEFINITION
AC025591      58431 bp      DNA      BCT      01-JUL-2000
AC025591      GI:8570488
VERSION
KEYWORDS       HTG.
SOURCE
ORGANISM       Staphylococcus aureus
                Staphylococcus aureus
                Bacillus/Clostridium group;
                Bacillus/Staphylococcus group; Staphylococcus.
                1 (bases 1 to 58431)
                Loh, P., Qi, S., Ray, L., Ford, B., Iondola, J. and Roe, B.A.
                Staphylococcus aureus BAC sabac-134
                Unpublished
                2 (bases 1 to 58431)
                Loh, P., Qi, S., Ray, L., Ford, B., Iondola, J. and Roe, B.A.
                Direct Submission
                Submitted (12-MAR-2000) Department Of Chemistry And Biochemistry,
                The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
                OK 73019, USA

```

On Jun 17, 2000 this sequence version replaced gi:8567825.

```

COMMENT
FEATURES
  source
    1. 58431
    /organism="Staphylococcus aureus"
    /db_xref="taxon:1280"
    /clone="sabac-134"
    /clone_lib="OU staph library"
BASE COUNT      19747 a      9170 c      10899 g      18615 t
ORIGIN

```

```

Query Match      7.4%; Score 40.4; DB 1; Length 58431;
Best Local Similarity 51.7%; Pred. No. 4.4;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

```

```

QY  91 gatgtcaaacctatgttggtgccaataatgtgcgaagtgaagccgaacaaatcaaggt 150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB  39398 GATGCTACGTATTCTATGTAAGACATGGGTACACTTTATACCAACACGTTGAAATTG 39457
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```
Oy 151 aagaacacccgtactatggtattatgcaggtataacttgacccaaatttggcgtagaa 210
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39458 GAAACCAAGGCGTGCACCTTTATGAAAGTATGAAAGTGAAGCTGAAGCAAGCA 39517
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 211 ccggaattgttggtcgaacgcacaaagaatttaacgacgcgtgagtcctgtaaaag 268
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 39518 AAGACTTGATTGATGACGACTTAATGATACTGAACCACTGTTGTGAACCGGACACAG 39575

RESULT 15
LOCUS AE001127/c
DEFINITION Borrelia burgdorferi (section 13 of 70) of the complete genome.
ACCESSION AE001127
VERSION AE001127.1 GI:2688047
KEYWORDS
SOURCE
ORGANISM
    Lyme disease spirochete.
    Borrelia burgdorferi
    Bacteria: Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
    burgdorferi group.
    1 (bases 1 to 10976)
REFERENCE
    Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
    Lachigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
    Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,
    Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
    Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,
    Gocayne,J.D., Weidman,J., Ullrich,T., Whiteley,L., McDonald,L.,
    Ariach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
    Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
    Direct Submission
    Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
    Medical Center Dr, Rockville, MD 20850, USA
TITLE
    Genomic sequence of a Lyme disease spirochete, Borrelia
    burgdorferi
JOURNAL
    Nature 390 (6660), 580-586 (1997)
MEDLINE
    98065943
REFERENCE
    2 (bases 1 to 10976)
AUTHORS
    Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.A.,
    Lachigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
    Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D.,
    Richardson,D., Peterson,J., Kerlavage,A.R., Quackenbush,J.,
    Salzberg,S., Hanson,M., van Vugt,R., Palmer,N., Adams,M.D.,
    Gocayne,J.D., Weidman,J., Ullrich,T., Whiteley,L., McDonald,L.,
    Ariach,P., Bowman,C., Garland,S., Fujii,C., Cotton,M.D., Horst,K.,
    Roberts,K., Hatch,B., Smith,H.O. and Venter,J.C.
TITLE
    Submitted (12-DEC-1997) The Institute for Genomic Research, 9712
JOURNAL
    Medical Center Dr, Rockville, MD 20850, USA
FEATURES
    source
        location/Qualifiers
            1..10976
            /organism="Borrelia burgdorferi"
            /db_xref="taxon:139"
            107..823
            /gene="BB0158"
            107..823
            /gene="BB0158"
            /note="similar to PID:1063419 GB:AE000790 percent
            identity: 37.28; identified by sequence similarity:
            putative"
            /codon_start=1
            /transl_table=11
            /product="antigen, S2, putative"
            /protein_id="AAC66553.1"
            /db_xref="GI:2688051"
            /translation="MVPRYKHLIMLPMILSCAPFKKPOSHODSNPKPLSDEK
            LHLISGKSNKKLPIINSNDHYTWIKRAMLIGEDGCEIPEFKNRPYSYIISYVK
            DGKSYASALILFETTRKNGDEYEIDVKTGTVTGSTIELKNSLAVNSOEGYVTA
            YPFGILMSDEIKNAFKLTKNGHWNMYLADLVTKNKLQETKRIKYSLSKLIIEFLK
            EVLEKNSLKDIDAGDFEDI"
            847..1521
            /gene="BB0159"
            847..1521
            /gene="BB0159"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BB0159"
            /db_xref="GI:2688056"
            /translation="MSIKKFIITLILSIKNSFSENEINENINENIYIKENIKTEIK
            LKROSELASDVAVISOPKPIELADNGPIKLEGISYFINVFSKISSALISDLS
            NEASKYKIKIEPLSPKGNFINJLSSITSKOSKSKLADAISEFTGLRTRESKST
            IAEYKDNMYIILAIIVENNINKEFEYELRINPKIYNDFOKRLRHFNSQIKKF
            PIPLE"
            1569..2687
            /gene="BB0160"
            1569..2687
            /gene="BB0160"
            /note="similar to SP:p29743 PID:366388 GB:U00096
            PID:1790487 percent identity: 29.89; identified by
            sequence similarity; putative"
            /codon_start=1
            /transl_table=11
            /product="alanine racemase (alt)"
            /protein_id="AAC66549.1"
            /db_xref="GI:2688050"
            /translation="MYNNKTWGSNKTIIINLNLEHNLIKNKIGKEIYATLKGA
            YGHGLINTEFLKSKINYPGLSNIEDAKTLKIDKNKILMYIKVDKKEIKNLKFE
            LVPEYDFEYFLIEKECALOKNKIKVHLKIDGNRRGKIKDLEIATYIONSKFL
            LEEGICSHLPSTENKTKOKOIEOPLEFLTKOKNIHPRFVHISNGHIINYKINPO
            FMNVRPGLIYGCOSLKNKKPALNFKRYLSLFSKVIYIKNKKGEKTSYSGIPAKK
            DMKIGIIPYGFDPQNI SNDFYFLINNNKKGKIGKVCMLTIVETFRDLKVKTSK
            VEIYSEKLSIDMSKFSKRSHELLCNIGYKKRYLD"
            complement(2679..4343)
            /gene="BB0161"
            complement(2679..4343)
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BB0161"
            /protein_id="AAC66554.1"
            /db_xref="GI:2688055"
            /translation="MIDLKNTYONRVDSLRLMLFIIFISAPLSKEYYEYRPA
            DSVNSIVIGKEITKEESEEEYKFTDRNNNLIVYIGLSLIMPSFMAOSQIKI
            ERSKRSERIRNGLAFKNNNGVIEIIEYSDGRINIEFVNSNMPVRKQVSY
            YHFLVDNREFSVYFNSGQIKDNNVYEFKISYDTRKFKVTVLYDENGIVYRSK
            NGTEGFRLLYDNNHLIKEEYLDKSAVYNSPVSATYIHYDNGCKYIELINYYIN
            NLTPDENVALIYREYTFKDKDCYIKFENYDNNNLJTSONYAKKTFIYONNEKR
            IINYSNKTINKYNDIPTYEKEIKMNFAGIYSKYDNEFYLIENIFDKNPNL
            IADYGVMIYRYSYDKGFLIAOEHFGDFVNPIDDFGVSKYKFSYSGNMGIMSKN
            YSKDGLVADCNLFVEYELVEYDKQRLLSOKNFGSLGQLODDIHGVSYKYEYNEFK
            ISKOSNHPDLRLRDVCGFSIYKWIYDSRGLIDYQKFDLSGLNI"
            4357..4524
            /gene="BB0162"
            4357..4524
            /gene="BB0162"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BB0162"
            /protein_id="AAC66553.1"
            /db_xref="GI:2688054"
            /translation="MAKISKNAQVGSKELISRMGCKIIMSKFENCKIKHYACQTS
            KNTAKRPDLP"
            complement(4551..6299)
            /gene="BB0163"
            complement(4551..6299)
            /gene="BB0163"
            /note="hypothetical protein; identified by Glimmer;
            putative"
            /codon_start=1
            /transl_table=11
            /product="B. burgdorferi predicted coding region BB0163"
            /protein_id="AAC66552.1"
            /db_xref="GI:2688053"
            /translation="MKRLPRTFYMSKSEVDILISOLGISSTOKSIITLISYKKEAI
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 5, 2001, 01:54:04 ; Search time 79.21 Seconds  
(without alignments)  
2575.241 Million cell updates/sec

Title: US-09-164-714-6  
Perfect score: 543  
Sequence: 1 atgaagaacttaaaacact.....gcgcatttgcttttaa 543

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues  
Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_36.\*  
1: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1980.DAT:\*  
2: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1981.DAT:\*  
3: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1982.DAT:\*  
4: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1983.DAT:\*  
5: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1984.DAT:\*  
6: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1985.DAT:\*  
7: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1986.DAT:\*  
8: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1987.DAT:\*  
9: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1988.DAT:\*  
10: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1989.DAT:\*  
11: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1990.DAT:\*  
12: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1991.DAT:\*  
13: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1992.DAT:\*  
14: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1993.DAT:\*  
15: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1994.DAT:\*  
16: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1995.DAT:\*  
17: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1996.DAT:\*  
18: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1997.DAT:\*  
19: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1998.DAT:\*  
20: /cgnl\_8/gcgdata/geneseq/geneseqn/NA1999.DAT:\*  
21: /cgnl\_8/gcgdata/geneseq/geneseqn/NA2000.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	543	100.0	543	21	DNA encoding the o
2	71.2	13.1	80	21	Nucleotide sequenc
3	41	7.6	1762	16	B31 Osp-A/antigen
4	40.4	7.4	18613	18	Staphylococcus aur
5	39.4	7.3	846	17	OspB soluble varia
6	39.4	7.3	891	16	B. burgdorferi str
7	39.4	7.3	891	17	OspB variant #1 co
8	39.4	7.3	1141	16	B31 outer surface
9	39.4	7.3	1180	16	B31 outer surface
10	39.4	7.3	1324	16	B31 outer surface
11	39.4	7.3	1363	16	B31 outer surface
12	39.4	7.3	1915	11	OspA and OspB-enco

13	39.4	7.3	1916	19	V00681	Borrelia burgdorfe
14	39.4	7.3	5898	19	X00677	DNA plasmid VR2211
15	39.4	7.3	53585	20	X20251	Borrelia burgdorfe
16	38.4	7.1	3323	18	T84223	DNA encoding a g1u
17	38.4	7.1	3323	19	V53546	DNA encoding 2 Stra
18	36.4	6.7	400	18	V78384	Staphylococcus aur
19	35.6	6.6	3794	19	V00680	Plasmid pRR46 con
20	35.4	6.5	1214	18	V74424	Staphylococcus aur
21	35	6.4	19446	19	V52184	Streptococcus pneu
22	34.8	6.4	45	21	A12599	PCR primer used to
23	34.6	6.4	1349	19	X14172	H. pylori GPO 639
24	34	6.3	4373	18	T72719	Pyruvate:Flavodoxi
25	33.4	6.2	4023	20	X13023	Enterococcus faeca
26	33.2	6.1	1015	14	O40695	B. burgdorferi str
27	32.4	6.0	2455	19	V26363	Moraxella catarrha
28	32.4	6.0	2896	13	O24135	Vector comprising
29	32	5.9	40	21	A12603	PCR primer used to
30	31.8	5.9	730	19	T98768	DNA encoding a S.
31	31.8	5.9	2531	19	V42988	Streptococcus jann
32	31.4	5.8	164976	19	V21209	Methanococcus jann
33	31.2	5.7	584	19	V62140	HSV-2 strain SB5 C
34	31	5.7	12438	20	X13110	Ap Serotype 7 60KD
35	31	5.7	235033	19	V57926	Enterococcus faeca
36	31	5.7	237326	19	V57903	Hereditary haemoch
37	30.8	5.7	2407	19	V26361	Moraxella catarrha
38	30.8	5.7	2520	21	Z38336	Moraxella catarrha
39	30.8	5.7	2520	21	Z35587	M. catarrhalis COP
40	30.8	5.7	2520	21	Z35588	M. catarrhalis COP
41	30.8	5.7	3762	18	T73217	AprICA gene. Acti
42	30.8	5.7	3762	18	T88584	A. pleuropneumonia
43	30.8	5.7	4042	16	T04132	Helicobacter pylor
44	30.8	5.7	7528	20	X12992	Enterococcus faeca
45	30.8	5.7				

#### ALIGNMENTS

RESULT 1  
ID A12591  
AL2591 standard; DNA; 543 bp.  
XX AC A12591;  
XX DT 25-JUL-2000 (first entry)  
DE DNA encoding the outer membrane protein 21 of strain ATCC49143.  
XX XX  
KW Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;  
KW ctitis media; respiratory infection; sinusitis; pneumonia; immunisation;  
KW ds.  
XX OS Moraxella catarrhalis.  
XX FH Key  
XX FT 1..543 location/Qualifiers  
XX FT CDS  
XX FT /\*tag= a /product= "outer membrane protein 21"  
XX FT /transl\_except= (pos: 211..213, aa: Ale)  
XX PN WO200018910-A1.  
XX PD 06-APR-2000.  
XX PF 01-OCT-1999; 99WO-US22918.  
XX PR 01-OCT-1998; 98US-0164714.  
XX PA (ANTE-) ANTEX BIOLOGICS INC.  
XX PI Tucker K, Tillmann UF;  
XX DR WPI; 2000-293149/25.

DR P-PSDB: Y84612.

XX

PT Isolated outer membrane protein from a *Moraxella catarrhalis* strain

PT used for diagnosis, treatment and prevention of disease caused by *M.*

PT *catarrhalis* e.g. pneumonia, otitis media and respiratory infections -

XX

PS

XX Claim 11; Fig 3; 108pp; English.

XX

CC The present sequence encodes an outer membrane protein 21 (OMP21) of

CC *Moraxella catarrhalis* strain ATCC99143. The OMP21 protein has an

CC apparent molecular weight of 16-20 kD as determined by sodium

CC dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21,

CC its nucleic acids and antibodies can be used in prophylactic and

CC therapeutic compositions for treating a *M. catarrhalis* bacterial

CC infection, otitis media, respiratory infections, sinusitis and

CC pneumonia. They are useful as reagents for the clinical or medical

CC diagnosis of *M. catarrhalis* infections and for scientific research on

CC the properties of pathogenicity, virulence and infectivity of

CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly

CC those that are cytotoxic may be used in passive immunisation to prevent

CC or attenuate *M. catarrhalis* infections of animals e.g. humans.

XX

Sequence 543 BP; 165 A; 108 C; 118 G; 152 T; 0 other;

Query Match	100.0%;	Score 543;	DB 21;	Length 543;
Best Local Similarity	100.0%;	Pred. No. 4,8e-155;		
Matches 543; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY 1 atgaaaccttaaaacacacatatttggcagatcacgctcttcglttatatggcagatgagtcgt 60

Db 1 atgaaacctttaaacaacacatatttggcagatcacgctcttcglttatatggcagatgagtcgt 60

QY 61 aacgcctgcacatcaagcclatgagcaaltctcgtcgtalgcctcaaccctatglttggtccaaat 120  
|||||  
Db 61 aacgcctgcacatcaagcclatgagcaaltctcgtcgtalgcctcaaccctatglttggtccaaat 120  
|||||

QY 121 ggtcaagttagacgcgcaaatcaacggttaagaacacccgttatgtattatgcaggt 180  
|||||  
Db 121 ggtcaagttagacgcgcaaatcaacggttaagaacacccgttatgtattatgcaggt 180  
|||||

QY 181 tataactttgacccaataatttcgycgttagaacccgaattgttcgttcagacgccaaga 240

Db 181 tataactttgacccaataatttcgycgttagaacccgaattgttcgttcagacgccaaga 240

QY	241	tttaatgcagagcgtagaglcctgttaaaggtgagtccttctgtgtcttaagcaca	300
Db	241	tttaatgcagagcgtagagtcctgttaaaggtgagtccttctgtgtcttaagcaca	300

QY	301	301	301
Db	301	301	301

Oy 361 actaaagtagatglttaccagccgtaatgcacactacatcactcaacaacaagcgacaaacc 420  
 |||||  
 Db 361 actaaatagatggttaccagccgtaatgcacactatccttcttcaacacaaaacccctcaaac 420  
 |||||

Oy	421	agccacagcgcgtgttgcttgctttaaaccatlagcaaatgtaggcgttgaagca	480
Db	421	agccctacagcgcgtgttgcttgctttaaaccatlagcaaatgtaggcgttgaagca	480

Oy 481 agctacactatcctarcagaagatgcgaatgcatgttggcgccatttgcttt 540  
|||||  
481 agctacactatcctarcagaagatgcgaatgcatgttggcgccatttgcttt 540

Oy	541	taa	543
541	lll		
543	lll	543	

RESULT	2
A12605	
ID	A12605 standard; DNA; 80 BP.

XX A12605;  
AC  
XX 25-JUL-2000 (first entry)  
DT  
XX  
DE Nucleotide sequence of a fragment of outer membrane protein 21 DNA.

KW Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;  
KW otitis media; respiratory infection; sinusitis; pneumonia; immunisation;  
KW ss.

OS	Moraxella catarrhalis.
XX	
PN	W0200018910-A1.

PD	06-APR-2000.
XX	
PF	01-OCT-1499; 99WO-US22918.

PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PR 01-OCT-1998; 98US-0164714.  
XX

aa Tucker K, Tillmann UF;  
PI  
XX  
DR WPI; 2000-293149/25.

PT Isolated outer membrane protein from a *Moraxella catarrhalis* strain used for diagnosis treatment and prevention of disease caused by *M. catarrhalis* e.g. pneumonia, otitis media and respiratory infections -

aa Example 8; Page 105; 108pp; English.  
ps  
xx  
cc The present sequence represents a PC

The present sequence represents a PCR amplified fragment of the outer membrane protein 21 (OMP21) gene from a *Moraxella catarrhalis* strain. The OMP21 protein has an apparent molecular weight of 16-20 kD as determined by sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in prophylactic and therapeutic compositions for treating a *M. catarrhalis* bacterial infection, otitis media, respiratory infections, sinusitis and pneumonia. They are useful as reagents for the clinical or medical diagnosis of *M. catarrhalis* infections and for scientific research on the properties of pathogenicity, virulence and infectivity of *M. catarrhalis* and host defence mechanisms. The antibodies, particularly those that are cytotoxic may be used in passive immunisation to prevent or attenuate *M. catarrhalis* infections of animals e.g. humans.

Sequence 80 BP; 26 A; 20 C; 18 G; 13 T; 3 other;

Query Match	13.18;	Score 71.2;	DB 21;	Length 80;
Best Local Similarity	91.28;	Pred. No. 1.2e-12;		
Matches 73; Conservative	2;	Mismatches 5;	Indels 0;	Gaps 0;

**OY**    91 gatgtcaaacctatgttgytgccaaattgltcagtagacgccaaacaatccaagt 150  
      ||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
**Db**     1 gaqcgscarcctgttgtgcgaatatlqtcaaqlaqacgccaaaccaatccaagt 60

```
QY      151 aagaacaccgctatgtat 170
          |||||
Db      61 aagaacaccgctacqqaat 80
```

RESULT	3
Q90739	

DT	30-JUL-1996	(first entry)
XX		
DE	B31	Osp-A/antigen p41 (122-234)/Osp-C fusion gene.
XX		

KW Strain; B31; antigen; antigenic domain; protein; Osp-C  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.

OS Borrelia burgdorferi;

XX Key Location/Qualifiers  
 FH Mat\_peptide 1..1762  
 FT /\*tag- a

XX WO9512676-A1.

XX 11-MAY-1995.

XX 27-OCT-1994; 94WO-US12352.

XX 29-APR-1994; 94US-0235836.

XX 01-NOV-1993; 93US-0148191.

XX (ASUV-) ASSOC UNIVERSITIES INC.

XX Dunn JJ, Luft BJ;

XX WPI: 1995-215034/28.

XX P-PSDB; R75746.

XX Chimeric protein comprising 2 or more antigenic Borrelia  
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno:diagnostic assays

XX Claim 43; Fig 37; 200pp; English.

CC The present sequence encodes a fusion protein comprising outer  
 CC surface protein A (Osp-A), antigen P41 (122-234) and Osp-C  
 CC sequences, from the B. burgdorferi strain B31. Using chemical or  
 CC enzymatic methods, peptide fragments of the parent proteins were  
 CC prep., and analysed by western blot to assess their ability to  
 CC bind different anti-Osp-A/P41/Osp-C monoclonal antibodies. The  
 CC information obt. was used to locate antigenic domains in the  
 CC proteins, the epitopes of which were mapped with the aid of site  
 CC directed mutagenesis. Identical analyses were performed on a  
 CC selection of antigens purified from a variety of B. burgdorferi  
 CC strains, the results from which were utilised in the prepn. of a  
 CC pool of antigenic Borrelia polypeptides, and corresponding  
 CC polynucleotides. Fusion proteins (i.e. the above protein)  
 CC comprising 2 or more antigenic Borrelia polypeptides, that do  
 CC not naturally occur in the same protein, can be used in the  
 CC treatment and diagnosis of Borrelia infections, i.e. as a  
 CC vaccine against Lyme borreliosis, in immunodiagnostic assays to  
 CC detect anti-Borrelia antibodies or to measure T-cell reactivity.

XX Sequence 1762 BP; 700 A; 291 C; 343 G; 428 T; 0 other;

Query Match 7.6%; Score 41; DB 16; Length 1762;  
 Best local Similarity 49.3%; Pred. No. 0.0072;

Matches 107; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 33 agcttctgtgtattgagatgagtgctaacgctgcacatgactatggcaattctgctga 92  
 DB 114 agactcagtgcttctgttattgtaataaatttctgttaagcaaaagaaataagctc 173  
 QY 93 tgcataccatcatgtgtgttccaaattgttcaagtgaagcagaagaataatcaaggtaa 152  
 DB 174 tggcaaatatgatattgaagaacaattgctcgggttggaacttaaggaacttcgataa 233  
 QY 153 gaacaccgcttatgtattatgcagggtataaacttgaccaaaatttggcgtagaacc 212  
 DB 234 aaacaaatggttcggaaccttggaaggttcaaaagccgtgacaagaagtataaataaac 293  
 QY 213 cgaattgtgtgttcagacgcgaagaatttaatgca 249

DB 294 agttctgtgatttaaacacagtaaccttaagaagca 330

RESULT 4

ID V74423 standard; DNA; 18613 BP.

AC V74423;

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #112.

KW Computer readable medium; vaccine; S.aureus infection; Immunodetection;  
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;  
 KW toxic shock syndrome; ds.

XX Staphylococcus aureus.

XX OS

FH Key Location/Qualifiers  
 FT misc\_feature 901..960

FT /\*tag- a

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 2701..2760

FT /\*tag- b

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 4501..4560

FT /\*tag- c

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 6301..6360

FT /\*tag- d

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 8101..8160

FT /\*tag- e

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 9901..9960

FT /\*tag- f

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 11701..11760

FT /\*tag- g

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT 13501..13560

FT /\*tag- h

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They  
 are included to maintain the nucleotide numbering  
 given in the specification for this DNA sequence"

FT misc\_feature 15301..15360

FT /\*tag- i

FT /\*note- "these bases represent a line of missing text in  
 the sequence listing in the specification. They

```
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
FT      misc-feature 17101..17160
FT      /*tag= */
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence"
XX      RP786519-A2.
XX      PD 30-JUL-1997.
XX      PE 07-JAN-1997; 97EP-0100117.
XX      PR 05-JAN-1996; 96US-0009861.
XX      PA (HUMA-) HUMAN GENOME SCI INC.
XX      PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX      PI Rosen CA;
XX      DR WPI; 1997-374922/35.
XX      PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
XX      PT stored on computer readable medium and used in the production of
XX      PT anti-S.aureus vaccines
XX      PS Claim 1; Page 641-651; 3271pp; English.
XX      CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
XX      CC of the invention. The DNA sequences are recorded on a computer readable
XX      CC medium, preferably selected from a floppy or hard disk, random access
XX      CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX      CC the S.aureus DNA sequences allows putative functions to be assigned so
XX      CC that protein-encoding or regulatory regions of commercial, therapeutic or
XX      CC industrial importance can be obtained. Specifically, sequences which are
XX      CC likely to encode antigens have been identified and these polypeptides can
XX      CC be used in a vaccine composition against S.aureus infection. The
XX      CC polypeptides can also be used in a kit for the immunodetection of
XX      CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
XX      CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX      CC skin and surgical wound infections, scalded skin syndrome, toxic shock
XX      CC syndrome, etc. Organisms transformed with the DNA sequences can be used
XX      CC for recombinant production of the polypeptides. The new DNA sequences
XX      CC (and their fragments) are useful as primers or probes for isolating
XX      CC homologues of any of the S.aureus DNA sequences contained on the
XX      CC computer readable medium.
XX      SQ Sequence 18613 BP; 5990 A; 2753 C; 3477 G; 5782 T; 611 other:
XX
Query Match 7.4%; Score 40.4; DB 18; Length 18613;
Best Local Similarity 51.7%; Pred. No. 0.03;
Matches 92; Conservative 0; Mismatches 86; Indels 0; Gaps 0;
QY 91 gatgtccaacctatgtgtgtccaaatgtgtcaagttagacgccaagaatcaacggt 150
DB 11111111111111111111111111111111111111111111111111111111
DB 5362 gatgtctcagctatcatgttgaagacatgtgtacacttataccaacaacgttgatcg 5421
QY 151 aagaacacgcgtatgtaattatgcaggtttaacttgaccacaatttggcgtaga 210
DB 1111111111111111111111111111111111111111111111111111111
DB 5422 gaaaaaagaagcgctgaccttattatgaagtaacgaagtgaagaagctgaagaagaaga 5481
QY 211 ccgaattgtgtgttcagacgccaagaattaatgcagcgctgtgctcgttaaaag 268
DB 1111111111111111111111111111111111111111111111111111111
DB 5482 aaaagcttgatgtatgacgttgtaataatgtatactgaaccgctgtgtaaacggacaag 5539
XX
RESUME 5
T43302
ID T43302 standard; DNA; 846 BP.
XX
```

```
AC      T43302;
XX
XX      DT 10-FEB-1997 (first entry)
XX
XX      DE OspB soluble variant coding sequence.
XX
XX      KW OspB; OspB; outer surface protein A; Borrelia; variable major protein 7;
XX      KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;
XX      KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;
XX      KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;
XX      KW ds.
XX      OS Borrelia burgdorferi.
XX
XX      PN US5571718-A.
XX
XX      PD 05-NOV-1996.
XX
XX      PE 08-SEP-1992; 92US-0941523.
XX      PR 08-SEP-1992; 92US-0941523.
XX      PR 21-DEC-1990; 90US-0632072.
XX
XX      PA (ASUT-) ASSOC UNIVERSITIES INC.
XX      PI Barbour AG, Dunn JJ;
XX      PI WPI; 1996-505409/50.
XX      DR P-PSDB; W08096.
XX      PT Soluble recombinant forms of Borrelia lipo:proteins - useful for
XX      PT vaccine prodn. for treatment of Lyme disease
XX
XX      PS Claim 4; Column 43-44; 49pp; English.
XX
XX      CC T43301-T43303 represent coding sequences for soluble recombinant Borrelia
XX      CC proteins of the invention. This sequence represents the coding sequence
XX      CC for the soluble recombinant variant of the Borrelia burgdorferi outer
XX      CC surface protein B (OspB). OspB is a surface lipoprotein of the B.
XX      CC burgdorferi spirochete. Borrelia spirochetes are responsible for a
XX      CC variety of human disorders including Lyme borreliosis, and relapsing
XX      CC fevers. The spirochete is transmitted to humans and animals through the
XX      CC bite of a tick, and can cause serious dermatological, arthritic,
XX      CC neurological and other pathogenic disorders in an infected host. This
XX      CC sequence is used to create recombinant host cells, and the encoded OspB
XX      CC protein can be isolated from the cytosol of one of these cells without
XX      CC the use of detergent. The encoded recombinant proteins can be used as
XX      CC antigens for the production of vaccines against Lyme disease. The
XX      CC recombinant proteins can also be used in immunoassays and other
XX      CC diagnostic screening methods to detect the presence of antibodies against
XX      CC Borrelia lipoproteins in the sera of infected patients.
XX      SQ Sequence 846 BP; 367 A; 134 C; 150 G; 195 T; 0 other:
XX
Query Match 7.3%; Score 39.4; DB 17; Length 846;
Best Local Similarity 48.8%; Pred. No. 0.016;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 33 agctctcgtatgtggcgtatgagtgatgcctgcacatgtaagcaattctgtcga 92
DB 1111111111111111111111111111111111111111111111111111111
DB 117 agactcagtgctctgttgaatgtaataaatttgglaagcaagaataaataagctc 17,
QY 93 tgcataccctatgttgtgtgccaataattgtcaagttagcgcgaagcaatacaagtgta 152
DB 1111111111111111111111111111111111111111111111111111111
DB 177 cggcaaatatgattlaagcgcaacaattgatcaagtlgtgaactttaaaggacttccgata 236
QY 153 gaacacgcgtatggtataltatgcaggtataacttgcaccaaaatttggcgtagaac 212
DB 1111111111111111111111111111111111111111111111111111111
DB 237 aaacaatgtcttcgaaccttgaaggttcaaaagcttgacgaagagtaagaattaaat 296
QY 213 cgaattgtgtgttcagacgccaagaatttaagca 249
XX
```

Db 297 agttctgctgatttaaacacagtaacctagaagca 333

RESULT 6  
ID Q90713 standard: DNA: 891 BP.  
XX Q90713;  
XX  
XX 31-JUL-1996 (first entry)  
XX  
XX B. burgdorferi strain B31 outer surface protein B (OspB-B31) DNA.  
DE  
XX Strain B31: outer surface protein; OspB; antigenic domain;  
KW chimeric protein; treatment; diagnosis; infection; vaccine;  
KW Lyme borreliosis; immunodiagnostic assay; antibody;  
KW T-cell reactivity; chimeric; ds.  
XX  
XX Borrelia burgdorferi.  
OS  
XX  
XX key Location/Qualifiers  
FH 1..891  
FT CDS /\*tag= a  
XX  
XX W09512676-A1.  
XX  
XX 11-MAY-1995.  
XX  
XX 27-OCT-1994; 94WO-US12352.  
XX  
XX 29-APR-1994; 94US-0235836.  
PR 01-NOV-1993; 93US-0148191.  
XX  
XX (ASU- ) ASSOC UNIVERSITIES INC.  
XX  
XX Dunn JJ, Luft BJ:  
XX  
XX WPI: 1995-215034/28.  
DR P-PSDB; R75726.  
XX  
XX Chimeric protein comprising 2 or more antigenic Borrelia  
PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
PT immuno:diagnostic assays  
XX  
XX Example 1; Fig 11; 200pp; English.  
XX  
XX The present sequence encodes the B. burgdorferi strain B31, outer  
CC surface protein B (OspB-B31), using chemical or enzymatic methods,  
CC peptide fragments of OspB-B31 were prep'd, and analysed by western  
CC blot to assess their ability to bind different anti-OspB monoclonal  
CC antibodies. The information obt'd. was used to locate antigenic  
CC domains in OspB-B31, the epitopes of which were mapped with the  
CC aid of site directed mutagenesis. Identical analyses were performed  
CC on a selection of Osp purified from a variety of B. burgdorferi  
CC strains, the results from which were utilised in the prepn. of a  
CC pool of antigenic Borrelia polypeptides, and corresponding  
CC polynucleotides. Chimeric proteins comprising 2 or more antigenic  
CC Borrelia polypeptides, that do not naturally occur in the same  
CC protein, can be used in the treatment and diagnosis of Borrelia  
CC infections, i.e. as a vaccine against Lyme borreliosis, in  
CC immunodiagnostic assays to detect anti-Borrelia antibodies or to  
CC measure T-cell reactivity.  
XX  
XX Sequence 891 BP; 380 A; 137 C; 160 G; 214 T; 0 other;

Query Match 7.3%; Score 39.4; DB 16; Length 891;  
Best Local Similarity 48.8%; Pred. No. 0.016;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

OY 33 agcttctgtatttgagcgatgactaacgcctgcacatgactatgcaattctgctga 92  
DB 162 agactcagtgcttctgttaatgtaataaatttgttaagcaagaagaaataagctc 221

OY 93 tgcacacccctatgttgcgcacaaatttgcacagtagacgcgaagaaacacaggttaa 152  
DB 222 cggcaaatatgatttaagagcaacattgcatcaggttgacttaaggaacttcggttaa 281  
OY 153 gaacaccgcttaatgattatgcatgaggtataacttgaccacaaatttggcgtgaacc 212  
DB 282 aaacaatgtctcgtgaacccttgaaggttcaaaagcgcgaagaggaagaaattaac 341  
OY 213 cgaatttgcgtcagacgcgaagaaatttaagca 249  
DB 342 agttctgctgatttaaacacagtaacctagaagca 378

RESULT 7  
ID T43315 standard: DNA: 891 BP.  
XX T43315;  
XX  
XX 10-FEB-1997 (first entry)  
XX  
XX OspB variant #1 coding sequence.  
XX  
XX  
XX OsPA: OspB; outer surface protein A; Borrelia; variable major protein 7;  
KW Borrelia hermslii; Vmp7; surface lipoprotein; spirochete; human; antigen;  
KW Lyme borreliosis; relapsing fever; dermatological disorder; Lyme disease;  
KW arthritic disorder; neurological disorder; vaccine; Borrelia lipoprotein;  
KW ds.  
XX  
XX Synthetic.  
OS  
XX  
XX US5571718-A.  
PN  
XX 05-NOV-1996.  
PD  
XX 08-SEP-1992; 92US-0941523.  
PF  
XX 08-SEP-1992; 92US-0941523.  
PR 21-DEC-1990; 90US-0632072.  
XX  
XX (ASU- ) ASSOC UNIVERSITIES INC.  
XX  
XX Barbour AG, Dunn JJ:  
XX  
XX WPI: 1996-505409/50.  
DR  
XX  
XX Soluble recombinant forms of Borrelia lipo:proteins - useful for  
PT vaccine prodn. for treatment of Lyme disease  
XX  
XX Example 8; Column 39-42; 49pp; English.  
XX  
XX This sequence represents the coding sequence for a soluble recombinant  
CC Borrelia outer surface protein B (OspB) variant proteins of the  
CC invention. OspB is a surface lipoprotein of the B. burgdorferi  
CC spirochete. Borrelia spirochetes are responsible for a variety of human  
CC disorders including Lyme borreliosis, and relapsing fevers. The  
CC spirochete is transmitted to humans and animals through the bite of a  
CC tick, and can cause serious dermatological, arthritic, neurological and  
CC other pathogenic disorders in an infected host. This sequence is used to  
CC create recombinant host cells, and the encoded OspB protein can be  
CC isolated from the cytosol of one of these cells without the use of  
CC detergent. The encoded recombinant proteins can be used as antigens for  
CC the production of vaccines against Lyme disease. The recombinant  
CC proteins can also be used in immunoassays and other diagnostic screening  
CC methods to detect the presence of antibodies against Borrelia  
CC lipoproteins in the sera of infected patients.  
XX  
XX Sequence 891 BP; 380 A; 137 C; 160 G; 214 T; 0 other;

Query Match 7.3%; Score 39.4; DB 17; Length 891;  
Best Local Similarity 48.8%; Pred. No. 0.016;

Matches	106;	Conservative	0;	Mismatches	111;	Indels	0;	Gaps
Qy	33	agctcttcgttatgagcgatgagtcacacgctgcccacgcatacgtacgtacgtcgtga	92					
Db	162	agactcagctgctcttgcttaatagtgaataaatttlltgtaagcaagaataaataagctc	22					
Oy	93	tgctcaaccctatgtctgtgcacaataatgtgtcaagtacgaagcacaataacgctaa	15					
Db	222	cggcaataatgatttaagaagacaacatcgatcaggtlgaacttaagaagactccgataa	28					
Oy	153	gaacacccctatggtatttaacgtacgtataactltgacccaataatttgcgtagaac	21					
Db	262	aaacacatggtcttgtaaaccccttgaaggttcacagccctgcagacagagtaaaataac	34					
Oy	213	cgaatttggltgcagacgccaagaatttaatgca	249					
Db	342	agttctgctgatttaaacacagtaaccttagaagca	378					
RESULT	8							
Q90737	Q90737	standard; DNA; 1141 BP.						
XX	XX	Q90737;						
XX	XX	30-JUN-1996 (first entry)						
XX	XX	B31 outer surface protein (Osp-A)/antigen P41 (140-234) fusion gene.						
XX	XX	Strain: B31; antigen; antigenic domain; protein;						
KW	KW	treatment; diagnosis; infection; vaccine; Lyme borreliosis;						
KW	KW	immunodiagnostic assay; antibody; T-cell reactivity;						
KW	KW	outer surface protein; Osp-A; antigen P41; fusion; ds.						
XX	XX	Borrelia burgdorferi.						
OS	OS							
XX	XX							
FH	FH	Key	Location/Qualifiers					
FT	FT	Mat_peptide	1..1141					
FT	FT	/tag= a						
XX	XX	W09512676-A1.						
XX	XX	11-MAY-1995.						
PD	PD							
XX	XX	27-OCT-1994;	94WO-US12352.					
PF	PF							
XX	XX	29-APR-1994;	94US-0235836.					
PR	PR	01-NOV-1993;	93US-0148191.					
XX	XX	(ASUV-) ASSOC UNIVERSITIES INC.						
PA	PA							
XX	XX	Dunn JJ, Luft BJ;						
PI	PI							
XX	XX	WPI, 1995-215034/28.						
DR	DR	P-PSDB; R75744.						
XX	XX							
PT	PT	Chimeric protein comprising 2 or more antigenic Borrelia						
PT	PT	polymeride(s) - useful in a vaccine against Lyme borreliosis and in						
PT	PT	immuno:diagnostic assays						
CC	CC	Claim 43; Fig 35; 200pp; English.						
XX	XX							
CC	CC	The present sequence encodes a fusion protein comprising outer						
CC	CC	surface protein A (Osp-A) and antigen P41 (140-234) sequences, from						
CC	CC	the B. burgdorferi strain B31. Using chemical or enzymatic methods,						
CC	CC	peptide fragments of the parent proteins were prep'd. and analysed						
CC	CC	by western blot to assess their ability to bind different						
CC	CC	anti-Osp-A/P41 monoclonal antibodies. The information obt. was used						
CC	CC	to locate antigenic domains in the proteins, the epitopes of which						
CC	CC	were mapped with the aid of site directed mutagenesis. Identical						
CC	CC	analyses were performed on a selection of antigens purified from a						
CC	CC	variety of B. burgdorferi strains, the results from which were						
CC	CC	utilised in the prepn. of a pool of antigenic Borrelia polyepitides,						

CC	and corresponding polynucleotides. Fusion proteins (i.e. the
CC	above protein) comprising 2 or more antigenic Borrelia polypeptides,
CC	that do not naturally occur in the same protein, can be used in the
CC	treatment and diagnosis of Borrelia infections, i.e. as a vaccine
CC	against Lyme borreliosis, in immunodiagnostic assays to detect
CC	anti-Borrelia antibodies or to measure T-cell reactivity.
SQ	Sequence 1141 BP; 449 A; 206 C; 221 G; 265 T; 0 other;
	Query Match           7.3%; Score 39.4; DB 16; Length 1141;
	Best Local Similarity   48.8%; Pred. No. 0.018;
	Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
OY	33 agcttccttgattatggcgtatgagtcgaacgtccatcacactatgccaatttcgcgca 92
DB	114 gaaccacagtgtccttctgtttaatgatgaataaaatlittttaaagcaagaanaaalacgctc 173
OY	93 tgcacaacctatgttgtgtgccaaaattggtcaagtagacgcgaacgaatcaacgtcaa 152
DB	174 cggcaaatatgatttaaggagcaaacatgatcatcggtlgaacttaagaagaacltcgataa 233
OY	153 gaacacgcctatggttatltatgcaggtttataacttlggaccaaaatltygcgtagaacc 212
DB	234 aaacaatggttcttgcgaaccttgaaagttccaagcctgcacaagtagtaaaattaac 293
OY	213 gcaatttgttgttcagacgcgaagaattatataca 249
DB	294 agttctcgtcatltaaacacagctaaccttagaagca 330
RESULT 9	
ID	Q90735 .
XX	Q90735 standard; DNA; 1180 BP.
XX	
AC	Q90735:
XX	
DT	30-JUL-1996 (first entry)
XX	
DE	B31 outer surface protein (Osp-A)/antigen P41 (122-234) fusion gene.
XX	
KM	Strain: B31; antigen: antigenic domain; protein:
KW	treatment; diagnosis: infection; vaccine: Lyme borreliosis;
KM	immunodiagnostic assay: antibody; T-cell reactivity:
OS	outer surface protein; Osp-A; antigen P41; fusion; ds.
XX	
XX	Borrelia burgdorferi.
FH	Key Location/Qualifiers
FT	Mat_peptide 1..1180
FT	/tag= a
XX	
PN	W09512676-A1.
XX	
PD	11-MAY-1995.
XX	
PF	27-OCT-1994; 94WO-US12352.
XX	
PR	29-APR-1994; 94US-0235836.
PR	01-NOV-1993; 93US-0148191.
XX	
PA	(ASU-) ASSOC UNIVERSITIES INC.
XX	
PI	Dunn JT, Luft BJ;
XX	
DR	WPJ; 1995-215034/28.
DR	P-PSDB; R75742.
XX	
PT	Chimeric protein comprising 2 or more antigenic Borrelia
PT	polypeptide(s) - useful in a vaccine against Lyme borreliosis and in
XX	immuno-diagnostic assays
XX	
CS	Claim 43: Fig 33; 200BP; English.



XX The present sequence encodes a fusion protein comprising outer  
 CC surface protein A (Osp-A) and antigen P41 (122-234) sequences, from  
 CC the B. burgdorferi strain B31. Using chemical or enzymatic methods,  
 CC peptide fragments of the parent proteins were prep'd., and analysed  
 CC by western blot to assess their ability to bind different  
 CC anti-Osp-A/P41 monoclonal antibodies. The information obtd. was used  
 CC to locate antigenic domains in the proteins, the epitopes of which  
 CC were mapped with the aid of site directed mutagenesis. Identical  
 CC analyses were performed on a selection of antigens purified from a  
 CC variety of B. burgdorferi strains, the results from which were  
 CC utilised in the prep'n. of a pool of antigenic Borrelia polypeptides,  
 CC and corresponding polynucleotides. Fusion proteins (i.e. the  
 CC above protein) comprising 2 or more antigenic Borrelia polypeptides,  
 CC that do not naturally occur in the same protein, can be used in the  
 CC treatment and diagnosis of Borrelia infections, i.e. as a vaccine  
 CC against Lyme borreliosis, in immunodiagnostic assays to detect  
 CC anti-Borrelia antibodies or to measure T-cell reactivity.

PA (ASOY-) ASSOC UNIVERSITIES, INC.  
 XX  
 XX Dunn JJ, Luft BJ.  
 XX  
 XX WPT: 1995-215034/28.  
 DR P-PSDB: R75745.  
 XX  
 XX Chimeric protein comprising 2 or more antigenic Borrelia  
 PT polypeptide(s) - useful in a vaccine against Lyme borreliosis and in  
 PT immuno-diagnostic assays  
 XX  
 PS Claim 43; Fig 36; 200pp; English.

Query Match 7.3%; Score 39.4; DB 16; Length 1180;  
 Best Local Similarity 48.8%; Pred. No. 0.019; Indels 0; Gaps 0;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

The present sequence encodes a fusion protein comprising outer  
 surface protein A (Osp-A) and antigen P41 (140-295) sequences, from  
 the B. burgdorferi strain B31. Using chemical or enzymatic methods,  
 peptide fragments of the parent proteins were prep'd., and analysed  
 by western blot to assess their ability to bind different  
 anti-Osp-A/P41 monoclonal antibodies. The information obtd. was used  
 to locate antigenic domains in the proteins, the epitopes of which  
 were mapped with the aid of site directed mutagenesis. Identical  
 analyses were performed on a selection of antigens purified from a  
 variety of B. burgdorferi strains, the results from which were  
 utilised in the prep'n. of a pool of antigenic Borrelia polypeptides,  
 and corresponding polynucleotides. Fusion proteins (i.e. the  
 above protein) comprising 2 or more antigenic Borrelia polypeptides,  
 that do not naturally occur in the same protein, can be used in the  
 treatment and diagnosis of Borrelia infections, i.e. as a vaccine  
 against Lyme borreliosis, in immunodiagnostic assays to detect  
 anti-Borrelia antibodies or to measure T-cell reactivity.

Qy 33 agctcttcgtattgagcagatgagcgtacacgtccatcagctatgcaattctgtga 92  
 Db 114 agaccacagctcttctgttaatgataaataatttttgaaagcaagaataaataatgctc 173  
 Qy 93 lgtcaaccctatgltgtgccaanaattgttcaagtagaagcgaacaaatcaacggtaa 152  
 Db 174 cggcaaatatgatttaagagcaacatgatcaggttgtaacttaaggaacttcgataa 233  
 Qy 153 gaacacccgtttagtattatgacgttatacttgaccacaaatttgggtgaacc 212  
 Db 234 aaacacatggtcttgcagaccccttgaaggttcaagcctgacgaagtaaaatlaaac 293  
 Qy 213 cgaattgtgttcagacgcgaagaatttaatga 249  
 Db 294 agttctgtctgatttaaacacagtaacctagaagca 330

Query Match 7.3%; Score 39.4; DB 16; Length 1324;  
 Best Local Similarity 48.8%; Pred. No. 0.02; Indels 0; Gaps 0;  
 Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

RESULT 10  
 Q90738 standard; DNA: 1324 BP.  
 AC Q90738;  
 XX 30-JUL-1996 (first entry)  
 DE B31 outer surface protein (Osp-A)/antigen P41 (140-295) fusion gene.  
 KW Strain; B31; antigen; antigenic domain; protein;  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.  
 OS Borrelia burgdorferi.

RESULT 11  
 Q90736 standard; DNA: 1363 BP.  
 AC Q90736;  
 XX 30-JUL-1996 (first entry)  
 DE B31 outer surface protein (Osp-A)/antigen P41 (122-295) fusion gene.  
 KW Strain; B31; antigen; antigenic domain; protein;  
 KW treatment; diagnosis; infection; vaccine; Lyme borreliosis;  
 KW immunodiagnostic assay; antibody; T-cell reactivity;  
 KW outer surface protein; Osp-A; antigen P41; fusion; ds.  
 OS Borrelia burgdorferi.

XX Key Location/Qualifiers  
 FT Mat\_peptide 1..1324  
 FT /\*tag= a  
 PN W09512676-A1.  
 XX 11-MAY-1995.  
 PD 27-OCT-1994; 94WO-US12352.  
 PF 29-APR-1994; 94US-0235836.  
 PR 01-NOV-1993; 93US-0148191.  
 XX

XX Key Location/Qualifiers  
 FT Mat\_peptide 1..1324  
 FT /\*tag= a  
 PN W09512676-A1.  
 XX 11-MAY-1995.  
 PD 27-OCT-1994; 94WO-US12352.  
 PF 29-APR-1994; 94US-0235836.  
 PR 01-NOV-1993; 93US-0148191.  
 XX

```

FM Mat_peptide 1..1363
FT /*tag= a
XX
XX WO9512676-A1.
XX
XX 11-MAY-1995.
XX
XX 27-OCT-1994; 94WO-US12352.
XX
XX 29-APR-1994; 94US-0235836.
XX 01-NOV-1993; 93US-0148191.
XX
XX (ASUY-) ASSOC UNIVERSITIES INC.
XX
XX Dunn JT, Luft BJ;
XX
XX WPI: 1995-215034/28.
XX
XX P-USDB; R75743.
XX
XX Chimeric protein comprising 2 or more antigenic Borrelia
XX polypeptide(s) - useful in a vaccine against Lyme borreliosis and in
XX immuno:diagnostic assays
XX
XX Claim 43; Fig 34; 200pp; English.
XX
XX The present sequence encodes a fusion protein comprising outer
XX surface protein A (Osp-A) and antigen P41 (122-295) sequences, from
XX the B. burgdorferi strain B31. Using chemical or enzymatic methods,
XX peptide fragments of the parent proteins were prep'd., and analysed
XX by western blot to assess their ability to bind different
XX anti-Osp-A/P41 monoclonal antibodies. The information obt'd. was used
XX to locate antigenic domains in the proteins, the epitopes of which
XX were mapped with the aid of site directed mutagenesis. Identical
XX analyses were performed on a selection of antigens purified from a
XX variety of B. burgdorferi strains, the results from which were
XX utilised in the prepn. of a pool of antigenic Borrelia polypeptides,
XX and corresponding polynucleotides. Fusion proteins (i.e. the
XX above protein) comprising 2 or more antigenic Borrelia polypeptides,
XX that do not naturally occur in the same protein, can be used in the
XX treatment and diagnosis of Borrelia infections, i.e. as a vaccine
XX against Lyme borreliosis, in immunodiagnostic assays to detect
XX anti-Borrelia antibodies or to measure T-cell reactivity.
XX
XX Sequence 1363 BP; 538 A; 241 C; 252 G; 332 T; 0 other:
SQ
Query Match 7.3%; Score 39.4; DB 16; Length 1363;
Best Local Similarity 48.8%; Pred. No. 0.02;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
OY 33 agcttcctgattatggcagatgactaagcgtgcacatgacatgacatgacatctgctga 92
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
114 agactcagcgtcttctgttaataaataatlltgaagcaagaanaaaatgagtc 173
OY 93 tgcataacctatgtgtgtgccaataatggtcaatgtagcgcgaagcaataacggtta 152
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
174 cggcaaatatgatttaagaagcaacatlgatcaggttgaacttaagaagcaactccgata 233
OY 153 gaaccacccctatggtattatgacagatlaaacttaccacaaatlltgggtgaacc 212
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
234 aaacaatggttctggaacctggaaggttcaagcctgacagagtaaaataaataac 293
OY 213 cgaatttltggtlcaagcgcgaagaatlttaatga 249
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11
294 agttctgtgatttaaacacagtaaccttaagaagca 330

```

```

DT 02-OCT-1990 (first entry)
XX
XX OspA and OspB-encoding sequence of Borrelia burgdorferi.
XX
XX Borrelia burgdorferi; Lyme disease; active immunisation;
XX passive immunisation; OspA; OspB; ss.
XX
XX Borrelia burgdorferi.
XX
XX Key Location/Qualifiers
XX RBS 1..138
XX /*tag= a
XX /*label= OspA RBS
XX CDS 151..972
XX /*tag= b
XX /*label= OspA structural gene
XX 102..107
XX /*tag= c
XX /*label= -10 region of p1
XX 108..113
XX /*tag= d
XX /*label= -10 region of p2
XX 80..85
XX /*tag= e
XX /*label= -35 region of p1
XX -35_signal 87..92
XX /*tag= f
XX /*label= -35 region of p2
XX RBS 972..977
XX /*tag= g
XX /*label= OspB RBS
XX CDS 982..1873
XX /*tag= h
XX /*label= OspB structural gene
XX
XX WO9004411-A.
XX
XX 03-MAY-1990.
XX
XX 24-OCT-1989; 89WO-WO00248.
XX
XX 24-OCT-1988; 88DK-0005902.
XX
XX (SYMB-) SYMBICOM AB.
XX
XX Bergstrom S, Barbour AG, Magnarelli LA;
XX
XX WPI: 1990-163873/21.
XX P-USDB; R05028 ,R05029
XX
XX New Immunologically active fractions of Borrelli burgdorferi -
XX and derived antibodies and DNA coding sequences for active and
XX passive immunisation against and diagnosis of Lyme disease.
XX
XX Disclosure: ; 112pp; English.
XX
XX OspA and OspB genes are part of the same operon located on a linear
XX plasmid of B. burgdorferi and encode outer membrane proteins. They
XX are separated by 12 bases within which is located the RBS of OspB.
XX There are two putative promoters (P1 and P2) upstream of OspA. They
XX include two closely spaced direct repeats of a 12-mer sequence
XX (start-13 and 29) and a 14-mer palindromic sequence (start-123) which
XX surrounds the P1 and P2 -10 regions.
XX
XX Sequence 1915 BP; 811 A; 279 C; 337 G; 488 T; 0 other:
SQ
Query Match 7.3%; Score 39.4; DB 11; Length 1915;
Best Local Similarity 48.8%; Pred. No. 0.023;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
OY 33 agcttcctgattatggcagatgactaagcgtgcacatgacatgacatctgctga 92
DB 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11 11

```

```

Db 1143 agactcagtcgtcttggtaattgtaataaatttltgttaagcaagaagaaaaatagctc 1202
Oy 93 tgcacaacctatgttggcgcaaaattgtgtcaagtagagcgcaagcaaatcaacggttaa 152
Db 1203 cggcaaatgtatgttaagagcaacaattatcaggttgaacttaaggaactccgataa 1262
Oy 153 gaacaccgtatgagtatattatgcaggttataaacttgaaccaaaatttggcgtagaacc 212
Db 1263 aaacaatggttcgcgaacccttgaaggttcaaaagcttgacaagaagtataaattaac 1322
Oy 213 cgaattgttgcgtcagcagcgcaagaatttaatgca 249
Db 1323 agttctgcgtatttaaacacagtaacttagaagca 1359

RESULT 13
V00681
ID V00681 standard; DNA; 1916 BP.
XX
AC V00681;
XX
DE 08-JUN-1998 (first entry)
XX
DE Borrelia burgdorferi ospA and ospB gene region.
XX
KM Borrelia: antigen; outer surface protein; OspA; OspB; Lyme disease;
KW DNA vaccine; immunisation; VR2210; VR2211; vector; ss.
XX
OS Borrelia burgdorferi strain B31.
XX
XX Key Location/Qualifiers
FH RBS 139..144
FT CDS /*tag= a
FT CDS 152..973
FT /*tag= b
FT sig_peptide 152..199
FT /*tag= c
FT mat_peptide 200..970
FT /*tag= d
FT RBS 973..979
FT /*tag= e
FT CDS 983..1873
FT /*tag= f
FT sig_peptide 983..1027
FT /*tag= g
FT mat_peptide 1028..1870
FT /*tag= h
XX
XX WO9747197-A1.
XX
XX 18-DEC-1997.
XX
XX 03-JUN-1997; 97WO-US09439.
XX
XX 14-JUN-1996; 96US-0663998.
XX
XX (INMR ) PASTEUR MERIEUX SERUMS & VACCINS.
XX (UYTE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
XX (VICA-) VICAL INC.
XX
XX Barbour AG, Garner KR, Huebner RC, Liang X, Luke CJ,
XX Norman JA;
XX
XX WPI: 1998-051938/05.
XX DR P-PSDB; W26771 AND W37131.
XX
XX Plasmid for expressing Borrelia antigen in eukaryotic cells - used
XX PT as vaccines to protect against Lyme disease and for production of
XX PT antigens, themselves used in vaccines
XX
XX Example 1; Fig 7A-H; 80pp; English.
XX
XX This nucleotide sequence comprises the ospA and ospB gene region

```

```

CC of Borrelia burgdorferi. The ospA and ospB genes respectively
CC (see W26771) of B. burgdorferi. The ospA and ospB genes have
CC been used in the construction of VR2210 (see V00676) and VR2211
CC (see V00677), DNA plasmids suitable for transfection of e.g. baby
CC hamster kidney and human melanoma cells. Plasmid DNA encoding at
CC least one Borrelia genospecies antigen is disclosed and claimed.
CC The genospecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or OspC.
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
XX
SQ Sequence 1916 BP; 810 A; 280 C; 337 G; 489 T; 0 other;

Query Match 7.3%; Score 39.4; DB 19; Length 1916;
Best Local Similarity 48.8%; Pred. No. 0.023;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Oy 33 agcttcttgattatggcgaagtgctaacgctgcacatcagctatgccaattcgtcga 92
Db 1144 agactcagtcgtcttggtaattgtaataaatttltgttaagcaagaagaaaaatagctc 1203
Oy 93 tgcacaacctatgttggcgcaaaattgtgtcaagtagagcgcaagcaaatcaacggttaa 152
Db 1204 cggcaaatgtatgttgaagcaacaattatcaggttgaacttaaggaactccgataa 1263
Oy 153 gaacaccgtatgagtatattatgcaggttataaacttgaaccaaaatttggcgtagaacc 212
Db 1264 aaacaatggttcgcgaacccttgaaggttcaaaagcttgacaagaagtataaattaac 1323
Oy 213 cgaattgttgcgtcagcagcgcaagaatttaatgca 249
Db 1324 agttctgcgtatttaaacacagtaacttagaagca 1360

RESULT 14
V00677
ID V00677 standard; DNA; 5898 BP.
XX
AC V00677;
XX
DE 08-JUN-1998 (first entry)
XX
XX DNA plasmid VR2211 containing Borrelia burgdorferi OspB gene.
XX
XX Borrelia: antigen; outer surface protein A; OspB; Lyme disease;
XX KW DNA vaccine; immunisation; VR2211; vector; ss.
XX
XX OS Chimeric - Borrelia burgdorferi strain B31.
XX
XX OS Chimeric - Homo sapiens.
XX
XX FH Key Location/Qualifiers
FT primer_bind /*tag= a
FT /*note= "Binding site of forward primer used to
FT amplify and insert OspB coding sequence"
FT primer_bind 835..852
FT /*tag= b
FT /*note= "Binding site of reverse primer used to
FT amplify and insert OspB coding sequence"
XX
XX WO9747197-A1.
XX
XX 18-DEC-1997.
XX
XX 03-JUN-1997; 97WO-US09439.
XX
XX 14-JUN-1996; 96US-0663998.

```

```

XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS.
PA (UTRE-) UNIV TEXAS HEALTH SCI CENT SAN ANTONIO.
PA (VICA-) VICAL INC.
XX
PI Barbour AG, Garner KR, Huebner RC, Llang X, Luke CJ;
PI Norman JA;
XX WPI: 1998-051938/05.
XX
DR WPI: 1998-051938/05.
XX
PT Plasmid for expressing Borrelia antigen in eukaryotic cells - used
PT as vaccines to protect against Lyme disease and for production of
PT antigens, themselves used in vaccines
XX
PS Example 1; Fig 4A-D; 80pp; English.
XX
XX This is the nucleotide sequence of VR2211, a DNA plasmid suitable
CC for transfection of e.g. baby hamster kidney and human melanoma
CC cells. VR2211 contains DNA encoding the Borrelia burgdorferi
CC outer surface protein B (OspB). It was made by ligating: an OspB
CC fragment (see V00680) amplified from pTRH46 (see V00685-86); a
CC tissue plasminogen activator 5'-untranslated region/leader peptide
CC sequence amplified from nKCMintBL (see V00678); and PstI- and
CC XbaI-digested plasmid VR1012. Plasmid DNA encoding at least one
CC Borrelia genospecies antigen is disclosed and claimed. The
CC genospecies may be B. burgdorferi, Borrelia garinii and/or
CC Borrelia afzelii. The antigen can be OspA and/or OspB and/or OspC.
CC The plasmid is used in vaccines to elicit a protective immune
CC response (both humoral and cellular) in hosts, human or animal,
CC susceptible to Lyme disease (claimed). It can also be used for the
CC production of a Borrelia antigen in vitro in eukaryotic cells
CC (claimed), and these antigens used as vaccines or immunogenic
CC compositions, or to produce monoclonal antibodies.
XX
SQ Sequence 5898 BP, 1643 A; 1402 C; 1380 G; 1473 T; 0 other;

Query Match          7.3%; Score 39.4; DB 19; Length 5898;
Best Local Similarity 48.8%; Pred. No. 0.037; Indels 0; Gaps 0;
Matches 106; Conservative 0; Mismatches 111;

QY 33 agctcttcgtatctggcgatgagtcgtacgcgtccatcagctatgcaattctgctga 92
DB 118 agactcagtgctctgttcttaagtcataataatttttgaagcaaaagaaaaataagctc 177
QY 93 tgcataacctatgttctgtgcgaataatggtcagtagcgcgaagcaaatcaacggtaa 152
DB 178 cggcaaatatgatttaagagcaacattgatacaggttgtaactlaaagaaacttcgataa 237
QY 153 gaacaccgctatggtatattatgcaggtataacttggcaccaaatttggcgtagaacc 212
DB 238 aaacaaatggtctctggaaccttgaaaggttcgaagccctgacagaglaaagaataaataac 297
QY 213 cgaatttgttgcagacgcgcaaaagatttaatgca 249
DB 298 agttctctgatttaaacacagtaaccttagaagca 334

RESULT 15
X20251/c
ID X20251 standard; DNA; 53585 BP.
XX
AC X20251;
XX
XX 04-MAY-1999 (first entry)
XX
DE Borrelia burgdorferi polynucleotide sequence #4.
XX
XX Borrelia burgdorferi; spirochete; bacterium; pathogen; Lyme disease;
KW epidemic relapsing fever; endemic relapsing fever; Lyme borreliosis;
KW infection; diagnosis; characterisation; detection; ds.
XX
OS Borrelia burgdorferi.

```

```

XX
PN WO9658943-A1.
XX
PD 30-DEC-1998.
XX
PF 18-JUN-1998; 98WO-US12764.
XX
PR 03-SEP-1997; 97US-0057483.
PR 20-JUN-1997; 97US-0050359.
PR 22-JUL-1997; 97US-0053344.
PR 22-JUL-1997; 97US-0053377.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (MEDI-) MEDIMUNE INC.
XX
PI Clayton R, Dougherty BA, Fraser C, Lathigra K, Smith HO;
PI White OR;
XX
DR WPI: 1999-081217/07.
XX
PT New isolated Borrelia burgdorferi nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of infections, particularly Lyme disease
XX
PS Claim 1; Page 801-831; 1128pp; English.
XX
XX X20248 to X20402 represent polynucleotide sequences isolated from
CC Borrelia burgdorferi (Bb). Products derived from Bb can be used for
CC the detection, diagnosis, characterisation, prevention and therapy of
CC Bb infections, e.g. Lyme disease. They can also be used for the
CC production of biosynthetic products, e.g. enzymes. Borrelia belongs
CC to a family of motile, spiral-shaped bacteria called Spirochetes.
CC Spirochetes are pathogenic in humans and Borrelia causes epidemic and
CC endemic relapsing fever, and Lyme borreliosis, more commonly known as
CC Lyme disease.
XX
SQ Sequence 53585 BP; 19235 A; 8081 C; 7001 G; 19266 T; 2 other;

Query Match          7.3%; Score 39.4; DB 20; Length 53585;
Best Local Similarity 48.8%; Pred. No. 0.096; Indels 0; Gaps 0;
Matches 106; Conservative 0; Mismatches 111;

QY 33 agctcttcgtatctggcgatgagtcgtacgcgtccatcagctatgcaattctgctga 92
DB 43176 AGACTCAGTCGTCCTTTGTTAATGCTAATTAATAATTTTGTAAAGCAAGAAAAATAGCTC 43117
QY 93 tgcataacctatgttctgtgcgaataatggtcagtagcgcgaagcaaatcaacggtaa 152
DB 43116 CCGCAAAATATGATTTTAAGACCAACATTTGATCAGTTGAACTTAAAGAACTTCGATAA 43057
QY 153 gaacaccgctatggtatattatgcaggtataacttggcaccaaatttggcgtagaacc 212
DB 43056 AAACAAATGGTTCTCGAACCCTTGAAAGTTCAAGCTGACCAAGATTAAGTAATAATTAAC 42997
QY 213 cgaatttgttgcagacgcgcaaaagatttaatgca 249
DB 42996 AGTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 42960

Search completed: March 5, 2001, 06:20:48
Job time: 16004 sec

```

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: March 5, 2001, 01:28:06 ; Search time 62.98 Seconds  
(without alignments)  
1389.489 Million cell updates/sec

Title: US-09-164-714-6

Perfect score: 543

Sequence: 1 atgaacttaaacact.....gcgtcattggttttaa 543

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database : Issued\_Patents\_NA:\*

- 1: /cgml\_7/ptodata/1/ina/5A.COMB.seq:\*
- 2: /cgml\_7/ptodata/1/ina/5B.COMB.seq:\*
- 3: /cgml\_7/ptodata/1/ina/6.COMB.seq:\*
- 4: /cgml\_7/ptodata/1/ina/6CTUS.COMB.seq:\*
- 5: /cgml\_7/ptodata/1/ina/6ackfilesl.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39.4	7.3	846	1	US-07-941-523-21
2	39.4	7.3	891	1	US-07-941-523-19
3	39.4	7.3	1959	1	US-08-137-175A-1
4	39.4	7.3	1959	3	US-08-479-017-1
5	39.4	7.3	5952	2	US-08-663-998-2
6	31	5.7	2696	1	US-07-961-522-1
7	31	5.7	2696	1	US-08-321-978-1
8	31	5.7	2696	1	US-08-321-978-1
9	31	5.7	2696	2	US-08-710-584-1
10	31	5.7	246240	2	US-08-724-394A-20
11	31	5.7	246240	2	US-08-724-394A-21
12	31	5.7	246240	2	US-08-724-394A-22
13	30.8	5.7	2520	2	US-08-450-351-1
14	30.8	5.7	2520	2	US-08-450-351-3
15	30.8	5.7	3762	3	US-08-772-270A-3
16	30.8	5.7	4042	4	PCT-US95-02219-1
17	30.8	5.7	4042	4	PCT-US95-02219-1
18	30.8	5.7	4042	4	PCT-US95-02219A-1
19	30.8	5.7	8370	2	US-08-488-706-1
20	30.6	5.6	2278	1	US-08-258-188-1
21	30.6	5.6	2278	1	US-08-526-813-1
22	30.6	5.6	2278	4	PCT-US95-08554-1
23	30.2	5.6	1033	3	US-08-961-083-211
24	30.2	5.6	152331	3	US-09-128-155-17
25	30.2	5.6	176373	3	US-09-128-155-17
26	29.8	5.5	1903	1	US-07-961-522-3
27	29.8	5.5	1903	1	US-08-217-438-3
28	29.8	5.5	1903	1	US-08-217-438-3

29	29.8	5.5	1903	1	US-08-321-978-3	Sequence 3, Appl1
30	29.8	5.5	1903	2	US-08-710-584-3	Sequence 3, Appl1
31	29.4	5.4	660	2	US-08-625-377-1	Sequence 1, Appl1
32	29.4	5.4	3960	3	US-08-470-260-2	Sequence 2, Appl1
33	29.4	5.4	3960	3	US-08-471-491-2	Sequence 2, Appl1
34	29.4	5.4	3960	3	US-08-466-662-2	Sequence 2, Appl1
35	29.2	5.4	839	3	US-08-642-807A-30	Sequence 30, Appl1
36	29.2	5.4	1529	2	US-08-477-451-17	Sequence 17, Appl1
37	29.2	5.4	1529	2	US-08-477-451-21	Sequence 21, Appl1
38	29.2	5.4	19932	2	US-08-477-451-25	Sequence 25, Appl1
39	29	5.3	4140	3	US-08-894-731-2	Sequence 2, Appl1
40	28.8	5.3	2949	3	US-08-433-522A-1	Sequence 1, Appl1
41	28.8	5.3	2949	3	US-09-135-166-1	Sequence 1, Appl1
42	28.8	5.3	3294	3	US-08-409-995-1	Sequence 1, Appl1
43	28.6	5.3	3294	3	US-08-685-467-1	Sequence 1, Appl1
44	28.6	5.3	5319	1	US-08-159-927-1	Sequence 1, Appl1
45	28.4	5.2	1716	3	US-09-029-267-1	Sequence 1, Appl1

## ALIGNMENTS

RESULT 1  
US-07-941-523-21  
; Sequence 21, Application US/07941523  
; Patent No. 5571718  
; GENERAL INFORMATION:  
; APPLICANT: Dunn, John J  
; APPLICANT: Barbour, Alan G  
; TITLE OF INVENTION: Cloning and Expression of Borrelia  
; NUMBER OF INVENTION: Lipoproteins  
; CORRESPONDENCE ADDRESS: 24  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 01730  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/941,523  
; FILING DATE: 19920908  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; TELEPHONE: (617) 861-6240  
; TELECOMMUNICATION INFORMATION:  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 846 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-941-523-21

Query Match 7.3%; Score 39.4; DB 1; Length 846;  
Best Local Similarity 48.8%; Pred. No. 0.0037;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Caps 0;

QY 33 agctcttcgtatctgcatgagtgctcaacgctgcacatcgcatatgcaattctgcga 92  
DB 117 AGACTCAGCTCTTGTGTAATGATTAATAATTTTGTATGACAAACAAAAAATACCTC 176



NAME/KEY: CDS  
LOCATION: 950..1840  
OTHER INFORMATION: /product= "ospb"  
US-08-137-175A-1

Query Match 7.3%; Score 39.4; DB 1; Length 1959;  
Best Local Similarity 48.8%; Pred. No. 0.0053;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Db 33 agctcttcgtatgtgcagatgagtgctaacgcgcacatcagctatgagcaattcgtcga 92  
1111 AGACTCAGTGTCTGTTTAAATGTAATTAATTTTGAAGCAAGAAAAAATAGCTC 1170  
Qy 93 tgcacacccatgltgltgcacaaattggtcagtagagcccaagcaatcagcgttaa 152  
1171 CGGCAATATGATTTAAGACACAAATGATCAGGTGAACCTTAAGAACTTCGATTA 1230  
Db 153 gaacacccatgltgltgacaggtatataacttgacaaaatttgcgtagaacc 212  
1231 AAGCAATGTTCTGACACCTTGAAAGGTTCAGACCTGACAGAGTAAGTAATTAAC 1290  
Qy 213 cgaatttgltgltcagagcccaagaattatgca 249  
Db 1291 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 1327

RESULT 4  
US-08-479-017-1  
Sequence 1, Application US/08479017  
Patent No. 6143872

GENERAL INFORMATION:  
APPLICANT: BARBOUR, Alan G.  
APPLICANT: BERGSTROM, Sven  
APPLICANT: HANSSON, Lennart  
TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND  
NUMBER OF SEQUENCES: 22  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BROMBY AND NEIMARK  
STREET: 419 Seventh Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,017  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/137,175  
FILING DATE: 26-OCT-1993  
APPLICATION NUMBER: PCT/US92/08972  
FILING DATE: 22-OCT-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: COOPER, Iver P.  
REGISTRATION NUMBER: 28,005  
REFERENCE/DOCKET NUMBER: BARBOUR-1B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-628-5197  
TELEFAX: 202-737-3528  
TELEX: 248633  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1959 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Borrelia burgdorferi  
STRAIN: B31 (ATCC 35210)

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 123..142  
OTHER INFORMATION: /function= "primer"

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 584..607  
OTHER INFORMATION: /function= "primer"

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 776..794  
OTHER INFORMATION: /function= "primer"

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 806..817  
OTHER INFORMATION: /function= "primer"

FEATURE:  
NAME/KEY: CDS  
LOCATION: 119..940  
OTHER INFORMATION: /product= "ospa"

FEATURE:  
NAME/KEY: CDS  
LOCATION: 950..1840  
OTHER INFORMATION: /product= "ospb"

US-08-479-017-1

Query Match 7.3%; Score 39.4; DB 3; Length 1959;  
Best Local Similarity 48.8%; Pred. No. 0.0053;  
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 33 agctcttcgtatgtgcagatgagtgctaacgcgcacatcagctatgagcaattcgtcga 92  
1111 AGACTCAGTGTCTGTTTAAATGTAATTAATTTTGAAGCAAGAAAAAATAGCTC 1170  
Db 93 tgcacacccatgltgltgcacaaattggtcagtagagcccaagcaatcagcgttaa 152  
1171 CGGCAATATGATTTAAGACACAAATGATCAGGTGAACCTTAAGAACTTCGATTA 1230  
Qy 153 gaacacccatgltgltgacaggtatataacttgacaaaatttgcgtagaacc 212  
1231 AAGCAATGTTCTGACACCTTGAAAGGTTCAGACCTGACAGAGTAAGTAATTAAC 1290  
Db 213 cgaatttgltgltcagagcccaagaattatgca 249  
1291 AGTTTCTGCTGATTTAAACACAGTAACCTTAGAAGCA 1327

RESULT 5  
US-08-663-998-2  
Sequence 2, Application US/08663998  
Patent No. 5845946

GENERAL INFORMATION:  
APPLICANT: HUEBNER, Robert C.  
APPLICANT: NORMAN, Jon A.  
APPLICANT: LIANG, Xiaowu  
APPLICANT: CARNER, Kristin R.  
APPLICANT: BARBOUR, Alan G.  
APPLICANT: LUKE, Catherine J.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,998
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454,312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-663-998-2

```

```

Query Match          7.3%; Score 39.4; DB 2; Length 5952;
Best Local Similarity 48.8%; Pred. No. 0.0087;
Matches 106; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```

```

QY 33 acgtcttcgtatagtgagtgatgacgttcacatcagctatgagcaatctgctga 92
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1980 AGAGCTAGAGTCTCTTTTAATGTAATAAATTTTGTAGCAAGAAAATAATAGCTC 2039
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 93 tgcacaacctatgttgcgcaaatgtgtcaagtagacgcgaagcaaatcaacgtaa 152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2040 CGGCAATATGATTTAAGAGCAACATGATGATGTTGAACCTTAAGGAACCTCGATTA 2099
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 153 gaaccacgctatagatattatgcaggtatataacttgcacaaatttggcgtagaac 212
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2100 AACAATGTTCTGGAACCTTGAAAGCTTCAAGCTGACAGAGTAAGTAATAATTAAAC 2159
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 213 cgaatttgtgttcagacgcgaagcaaatltaatga 249
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2160 AGTTTCTGCTGATTTAACACAGTAACCTTAGAGCA 2196
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 6
US-07-961-522-1
; Sequence 1, Application US/07961522
; Patent No. 5417971
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILLSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/961,522
; FILING DATE: 19921015

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0015.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2696 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 333..1973
; US-07-961-522-1

```

```

Query Match          5.7%; Score 31; DB 1; Length 2696;
Best Local Similarity 47.0%; Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;

```

```

QY 235 aaagaatttaagcagcgtagtccgtlaaaaagtatgtaagcttcttgcttat 294
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1662 ACAGAAATTGATGTTGATTTGGCAACAAAATTTAAGTGTAAATTTGATTAATAAAC 1721
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 295 ggcacatcgcataactcattcaatcccccttlaagcgaaggaatlaagcatt 354
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1722 GGCTGAATCTCTGTTTACCGTAGATG:MAAATGATGATGTAATGTTTACTGGCAAA 1781
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 355 gctaaagctaaagtagatgttaccagccgttaatgcaactacatcctcaaaaagcgac 414
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1782 GCTAANAACCTCAGATTAAGGCTTCGCTAGATTCAGGTAAGTTACAGTTATGAGATG 1841
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 415 aaacacagcctagcagcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1842 AATTTAAGAGTAGTACAGTTAGTGTGCTTATGTGTCAACGCGCA--CGAGAGCTT 1898
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 475 gaagcagctacacatcattcatagaagatgcgaatga 513
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1899 GCGGCAATTCACCATTAATCAGAAATGGCAGTGA 1937
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 7
US-08-217-438-1
; Sequence 1, Application US/08217438
; Patent No. 5521072
; GENERAL INFORMATION:
; APPLICANT: Potter, Andrew A.
; APPLICANT: Gerlach, Gerald F.
; APPLICANT: Willson, Philip J.
; APPLICANT: Rossi-Campos, Amalia
; TITLE OF INVENTION: ACTINOBACILLUS PLEUROPEINOMIAE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/217,438
; FILING DATE: 22-MAR-1994

```



```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-617-8999
TELEFAX: 415-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-217-438-1
```

```
Query Match          5.7%; Score 31; DB 1; Length 2696;
Best Local Similarity 47.0%; Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
```

```
QY 235 aaagaatttaatgcaaggcgtagtcctgtaaaagtgatgtaagctcttctggtgcttat 294
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1662 ACAGAAATTTGATGTTGATTTTGGCAACAAAATTAACTGTAAGTATTTGATAAANAAC 1721
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 ggcacatatgcctataaactcaatcaataaccatttaagcgaaggcgaataggacatt 354
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1722 GGGTGAATCTCTGTTTACCGTAGATGCAAAAATTGATGTAATGTTTACTGCGCAA 1781
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 gctaaagcctaagtagatglttacccagcgtaagcgaactacatactcaaaagaagcgac 414
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1782 GCTAAACCTCAGATGAGAGGCTTGCTCTAGATTCAAGTAGTTACGTTATGAGAAATGTG 1841
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 aaaaccagcctagcaggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1842 AAATTTAACGATGTAGACAGTTAGTGTGCTTCTATGTGTCCAAACGGCA---CCAGACCTT 1898
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 gaagcaagctacaactatctatcagaagaatgccaatgca 513
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1899 GCGGACAAATTCACCATTAATCAGAAATGCGAGTGTA 1937
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 8
US-08-321-978-1
; Sequence 1, Application US/08321978
; Patent No. 5801018
; GENERAL INFORMATION:
```

```
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; TITLE OF INVENTION: PLEUROPNEUMONIAE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 285 HAMILTON AVE, SUITE 200
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/321,978
; FILING DATE: 12-OCT-1994
```

```
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROBINS, ROBERTA L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 9001-0015.02
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 327-3400
TELEFAX: (415) 327-3231
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2696 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 333..1973
US-08-321-978-1
```

```
Query Match          5.7%; Score 31; DB 1; Length 2696;
Best Local Similarity 47.0%; Pred. No. 2.6;
Matches 131; Conservative 0; Mismatches 145; Indels 3; Gaps 1;
```

```
QY 235 aaagaatttaatgcaaggcgtagtcctgtaaaagtgatgtaagctcttctggtgcttat 294
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1662 ACAGAAATTTGATGTTGATTTTGGCAACAAAATTAACTGTAAGTATTTGATAAANAAC 1721
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 295 ggcacatatgcctataaactcaatcaataaccatttaagcgaaggcgaataggacatt 354
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1722 GGGTGAATCTCTGTTTACCGTAGATGCAAAAATTGATGTAATGTTTACTGCGCAA 1781
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 gctaaagcctaagtagatglttacccagcgtaagcgaactacatactcaaaagaagcgac 414
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1782 GCTAAACCTCAGATGAGAGGCTTGCTCTAGATTCAAGTAGTTACGTTATGAGAAATGTG 1841
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 aaaaccagcctagcaggcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 474
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1842 AAATTTAACGATGTAGACAGTTAGTGTGCTTCTATGTGTCCAAACGGCA---CCAGACCTT 1898
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 gaagcaagctacaactatctatcagaagaatgccaatgca 513
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1899 GCGGACAAATTCACCATTAATCAGAAATGCGAGTGTA 1937
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 9
US-08-710-584-1
; Sequence 1, Application US/08710584
; Patent No. 5876725
; GENERAL INFORMATION:
```

```
; APPLICANT: POTTER, ANDREW A.
; APPLICANT: GERLACH, GERALD F.
; APPLICANT: WILSON, PHILIP J.
; APPLICANT: ROSSI-CAMPOS, AMALIA
; TITLE OF INVENTION: VACCINES FOR ACTINOBACILLUS
; TITLE OF INVENTION: PLEUROPNEUMONIAE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBERTA L. ROBINS
; STREET: 285 HAMILTON AVE, SUITE 200
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/710,584
; FILING DATE:
```



APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONFIG"  
US-08-724-394A-21

Query Match 5.7%; Score 31; DB 2; Length 246240;  
Best Local Similarity 51.0%; Pred. No. 19;  
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 16 acacattgacagatcagctctctcgtatgtggcagatgagtgacgtccacacagc 75  
DB 75473 ACACACTTGCACTCAACAACTTCAGATTTATTCCTGCTGACTGACCTCCACATTATTC 75532  
QY 76 tatggcaattctgctatgctcaaccctatgttgytgccaaatgtgcaagtgaagcgc 135  
DB 75533 AGAGCCACACTTCCTTTATGGAATATTATAGCTTGTGCAAAAGTATTTCGGCTTTGCC 75592

QY 136 aagcaaatcaacgytaagaacac 158  
DB 75593 ATTAAAGTAAAGCAAAACACAC 75615

RESULT 12  
US-08-724-394A-22  
Sequence 22, Application US/08724394A  
Patent No. 5872237  
GENERAL INFORMATION:  
APPLICANT: Feder, John N.  
APPLICANT: Krommal, Gregory S.  
APPLICANT: Lauer, Peter M.  
APPLICANT: Ruddy, David A.  
APPLICANT: Thomas, Winston  
APPLICANT: Tsuchihashi, Zenta  
APPLICANT: Wolff, Roger K.  
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el  
TITLE OF INVENTION: Sequences and Antibodies There to  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,394A  
FILING DATE: 01-OCT-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:

NAME: Fitts, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 017957-000100  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 246240 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..246240  
OTHER INFORMATION: /note= "HLA-H.CONFIG"  
US-08-724-394A-22

Query Match 5.7%; Score 31; DB 2; Length 246240;  
Best Local Similarity 51.0%; Pred. No. 19;  
Matches 73; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 16 acacattgacagatcagctctctcgtatgtggcagatgagtgacgtccacacagc 75  
DB 75473 ACACACTTGCACTCAACAACTTCAGATTTATTCCTGCTGACTGACCTCCACATTATTC 75532

QY 76 tatggcaattctgctatgctcaaccctatgttgytgccaaatgtgcaagtgaagcgc 135  
DB 75533 AGAGCCACACTTCCTTTATGGAATATTATAGCTTGTGCAAAAGTATTTCGGCTTTGCC 75592

QY 136 aagcaaatcaacgytaagaacac 158  
DB 75593 ATTAAAGTAAAGCAAAACACAC 75615

RESULT 13  
US-08-450-351-1  
Sequence 1, Application US/08450351  
Patent No. 5981213  
GENERAL INFORMATION:  
APPLICANT: Hansen, Eric J.  
APPLICANT: Heilmann, Merja E.  
APPLICANT: Maciver, Isobel  
TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING TO  
TITLE OF INVENTION: USEFUL ANTIGENS OF MORAXELLA CATARRHALIS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,351  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: AMCY:019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:



```

; FILING DATE: December 23, 1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Gravelle, Michelle
; REGISTRATION NUMBER: 40,261
; REFERENCE/DOCKET NUMBER: 6580-81
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 364-7311
; TELEFAX: (416) 361-1398
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3762 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Actinobacillus pleuropneumoniae
; US-08-772-270A-3

```

```

Query Match          5.7%; Score 30.8; DB 3; Length 3762;
Best Local Similarity 49.4%; Pred. No. 3.5;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 87  tgcctgatgtcaacctatgttggcgcaaatgtgtcaagttagcgcgaagcaaatcaa 146
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2100  tgcctatgtacgcacacgctggcgatgcattgcggaacttgccgcgcatctactcgcaa 2159
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 147  cggtaagaacacgccttatgtgtattatgcaggttaacttggaccacaatttggcgt 206
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2160  aggttctgatgacgaaagccgcttaagctttagcttgcattgcttgaaagcgaagaaacctttt 2219
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY 207  agaaccgcaattgtgttcagacgcacaaagaatttaatgc 248
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 2220  agagaaagacccgcatcgttttgatataaaagtgctttgatcc 2261
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

Search completed: March 5, 2001, 06:21:36  
 Job time: 17610 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 4, 2001, 21:46:48 Search time 914.71 Seconds  
(without alignments)  
4159.860 Million cell updates/sec

Title: US-09-164-714-6  
Perfect score: 543  
Sequence: 1 atgaataacttaaaaacact.....gcgtccatttgcttttaa 543

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 7991742 seqs, 3503743858 residues

Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST\*

1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun1:\*  
43: em\_esthum1:\*

44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estp11:\*  
70: em\_estp12:\*  
71: em\_estp13:\*  
72: em\_estp14:\*  
73: em\_estp15:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est58:\*  
105: gb\_est59:\*  
106: gb\_est60:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estp16:\*  
 118: em\_estp17:\*  
 119: em\_estp18:\*  
 120: em\_estp19:\*  
 121: em\_estp20:\*  
 122: em\_estp21:\*  
 123: em\_estp22:\*  
 124: em\_estp23:\*  
 125: em\_estp24:\*  
 126: em\_estp25:\*  
 127: em\_estp26:\*  
 128: em\_estp27:\*  
 129: em\_estp28:\*  
 130: em\_estp29:\*  
 131: em\_estp30:\*  
 132: em\_estp31:\*  
 133: em\_estp32:\*  
 134: em\_estp33:\*  
 135: em\_estp34:\*  
 136: em\_estp35:\*  
 137: em\_estp36:\*  
 138: em\_estp37:\*  
 139: em\_estp38:\*  
 140: em\_estp39:\*  
 141: em\_estp40:\*  
 142: em\_estp41:\*  
 143: em\_estp42:\*  
 144: em\_estp43:\*  
 145: em\_estp44:\*  
 146: em\_estp45:\*  
 147: em\_estp46:\*  
 148: em\_estp47:\*  
 149: em\_estp48:\*  
 150: em\_estp49:\*  
 151: em\_estp50:\*  
 152: em\_estp51:\*  
 153: em\_estp52:\*  
 154: em\_estp53:\*  
 155: em\_estp54:\*  
 156: em\_estp55:\*  
 157: em\_estp56:\*  
 158: em\_estp57:\*  
 159: em\_estp58:\*  
 160: em\_estp59:\*  
 161: em\_estp60:\*  
 162: em\_estp61:\*  
 163: em\_estp62:\*  
 164: em\_estp63:\*  
 165: em\_estp64:\*  
 166: em\_estp65:\*  
 167: em\_estp66:\*  
 168: em\_estp67:\*  
 169: em\_estp68:\*  
 170: em\_estp69:\*  
 171: em\_estp70:\*  
 172: em\_estp71:\*  
 173: em\_estp72:\*  
 174: em\_estp73:\*  
 175: em\_estp74:\*  
 176: em\_estp75:\*  
 177: em\_estp76:\*  
 178: em\_estp77:\*  
 179: em\_estp78:\*  
 180: em\_estp79:\*  
 181: em\_estp80:\*  
 182: em\_estp81:\*  
 183: em\_estp82:\*  
 184: em\_estp83:\*  
 185: em\_estp84:\*  
 186: em\_estp85:\*  
 187: em\_estp86:\*  
 188: em\_estp87:\*  
 189: em\_estp88:\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	42	7.7	274	110	BE612078	BE612078 s103e08.y
2	39.4	7.3	1101	190	CNS0182P	AL108811 Drosophila
3	38.6	7.1	344	89	CNS34656	AW534656 S38H1 AGS
4	38.6	7.1	519	37	AV553491	AV553491 AV553491
5	38.6	7.1	1101	190	CNS00EAX	AV553491 AV553491
6	37.8	7.0	748	38	AV715018	AV715018 AV715018
7	37	6.8	740	38	AV679268	AV679268 AV679268
8	36.6	6.7	587	151	AO256569	AO256569 PC12B08 P
9	36.6	6.7	742	38	AV674403	AV674403 AV674403
10	36.6	6.7	479	171	AO826019	AO826019 HS_3442_A
11	36.4	6.7	481	29	AO835310	AO835310 AO835310
12	36.4	6.7	1101	190	CNS008P6	AL055205 Drosophila
13	36.4	6.6	446	94	AW575255	AW575255 s131h09.y
14	36	6.6	268	92	AW597222	AW597222 s171h02.y
15	35.8	6.6	410	95	AW781475	AW781475 s179d12.y
16	35.8	6.6	426	94	AW760255	AW760255 s148a03.y
17	35.8	6.6	435	94	AW733561	AW733561 s179d12.y
18	35.8	6.6	435	94	AW733561	AW733561 s179d12.y
19	35.8	6.6	454	94	AW757253	AW757253 s131h02.y
20	35.8	6.6	460	16	AL1138233	AL1138233 qd76e03.x
21	35.8	6.6	514	165	AO784640	AO784640 HS_3193_B
22	35.6	6.6	513	37	AV535288	AV535288 AV535288
23	35.6	6.6	549	94	AW754881	AW754881 PC06F10 P
24	35.6	6.6	614	191	CNS01WJQ	AL170495 Telradon
25	35.4	6.5	392	136	BE839457	BE839457 RC3-FN014
26	35.4	6.5	410	38	AV656509	AV656509 AV656509
27	35.4	6.5	591	94	AW755155	AW755155 PC11F05 P
28	35.4	6.5	624	97	AW981429	AW981429 EST92582
29	35.2	6.5	275	129	BB380142	BB380142 BB380142
30	35.2	6.5	807	190	CNS00A7D	AL054816 Drosophila
31	35	6.4	611	177	AZ225089	AZ225089 RPL1-23-6
32	35	6.4	1000	190	CNS00C0Q	AL059446 Drosophila
33	34.8	6.4	407	87	AW201163	AW201163 se98f03.y
34	34.8	6.4	556	150	AO176788	AO176788 HS_3213_A
35	34.8	6.4	1086	190	CNS00YXK	AL096962 Drosophila
36	34.6	6.4	353	147	Z46831	Z46831 ATTS4393 Ra
37	34.6	6.4	598	25	A1794544	A1794544 FC45f09.y
38	34.6	6.4	801	157	AO449605	AO449605 500002D09
39	34.4	6.3	385	93	AW620247	AW620247 s193f08.y
40	34.4	6.3	402	149	AO124099	AO124099 HS_3122_A
41	34.4	6.3	428	94	AW761487	AW761487 s166c04.y
42	34.4	6.3	430	94	AW757403	AW757403 s132h10.y
43	34.4	6.3	439	92	AW568592	AW568592 s178f12.y
44	34.4	6.3	442	87	AW202121	AW202121 s121b08.y
45	34.4	6.3	451	94	AW756981	AW756981 s128h03.y

## ALIGNMENTS

RESULT 1  
 LOCUS BE612078 274 bp mRNA  
 DEFINITION s103e08.y Gm-c1049 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1049-1599 5' similar to TR:080504 080504 F16H22.14 PROTEIN. ;,  
 mRNA sequence.  
 ACCESSION BE612078  
 VERSION BE612078.1 GI:9903110  
 KEYWORDS EST.





ACCESSION	sequence.
VERSION	AM334656
KEYWORDS	AM334656.1 GI:6831117
SOURCE	EST.
ORGANISM	Pneumocystis carinii f. sp. carinii. Pneumocystis carinii f. sp. carinii. Eukaryota; Fungi; Fungi Incertae sedis; Pneumocystidaceae;
REFERENCE	1 (bases 1 to 344) Smullen,A.G., Arnold,J., Weise,M., Wunderlich,J., Staben,C., Edman
AUTHORS	,J.C., Kovacs,J. and Cushion,M. Expressed sequence tags from Pneumocystis carinii
TITLE	Unpublished (2000)
JOURNAL	Contact: Staben C
COMMENT	School of Biological Sciences University of Kentucky 101 Morgan Building, University of Kentucky, Lexington, KY 40506-0225, USA Tel: 606 257 2161 Fax: 606 257 1717 Email: staben@pop.uky.edu. Location/Qualifiers
FEATURES	1..344
source	/db_xref="Pneumocystis carinii f. sp. carinii" /db_xref="taxon:38081" /clone_lib="AGS-1" /note="Vector: Lambda ZAP II; Site_1: EcoRI; Site_2: XhoI P. carinii organisms (3x10e9) from a single rat (99-1-6, sacrificed on 3/17/99) at Cincinnati VA facilities. Trisol extracted RNA. Oligo dT priming, standard conditions described by vendor, Stratagene. Further details see www.uky.edu/Project/Pneumocystis/"
BASE COUNT	99 a 43 c 50 g 152 t
ORIGIN	
Query Match	7.1%; Score 38.6; DB 89; Length 344;
Best Local Similarity	47.3%; Pctid No. 0.49;
Matches 116; Conservative 0; Mismatches 129; Indels 0; Gaps 0	
QY	127 gtagagcccaagcaatcaacgtagaagaccgcttagtattatgcaagtataac 186
Db	253 GAAAGCGCTTAATATTATTAAGAAATCTAATCTTATGAAGAACACGGGTATGTATA 194
QY	187 tttagcccaaatlttggcgtagaaccgcaattgttggtcgaagcccaagaattaat 246
Db	193 ATTGAACCAAAATCTTTGCTTTTTCGGGAACCGAACCTCATATTGCAAAAAATTATAT 134
QY	247 gcaagcgtgagtcctcgtaaaagtgatgtgaagcttlttggtcgtatgycacatctgc 306
Db	133 GTCGTAATCATGTGCGCAAAAGAGTAAATAATTCATTTTATCTCAGCATCTTTAGAC 74
QY	307 tataacttcaatatacccaattttagccaaagggcaaatltagcatlctaagactaa 366
Db	73 GAAATTCCTCTCAAAAAACCAAGTTATATATCAATGCAAGAGAGAAATTTTCAAAAAA 14
QY	367 gtaga 371
Db	13 AAAA 9
RESULT	4
LOCUS	AV553491
DEFINITION	AV553491 519 bp mRNA EST 06-SEP-2000
ACCESSION	AV553491 Arabidopsis thaliana roots Columbia Arabidopsis thaliana
VERSION	AV553491
KEYWORDS	AV553491.1 GI:8724904
SOURCE	EST.
ORGANISM	thale cress. Arabidopsis thaliana Eukaryote; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

REFERENCE	Magnoliophyta: eudicotyledons; core eudicots; Rosidae: eurosids II
AUTHORS	Brassicaceae: Brassicaceae: Arabidopsis.
TITLE	1 (bases 1 to 519) Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S. A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL	DNA Res. 7, 175-180 (2000)
MEDLINE	20363093
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yana 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: <a href="http://www.kazusa.or.jp/en/plant/">http://www.kazusa.or.jp/en/plant/</a> , Location/Qualifiers
FEATURES	Source 1..519 <sup>+</sup> /organism="Arabidopsis thaliana" <sup>+</sup> /strain="Columbia" <sup>+</sup> /db.xref="taxon:3702" <sup>+</sup> /clone="R263g09P" <sup>+</sup> /clone_1lb="Arabidopsis thaliana roots Columbia" <sup>+</sup> /tissue_type="roots" <sup>+</sup> /note="Vector: pluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT	144 a     90 c     122 g     163 t
ORIGIN	
Query Match	7.1%; Score 38.6; DB 37; Length 519;
Best Local Similarity	51.4%; Pred. No. 0.55;
Matches	89; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
Oy	122 gtcaagtagacgcgaagaatcgaactaacaggtaagaaccgcgtltagtalatllatgcaggt 181 
Db	326 GTTAGTGTTGACAGCACTGTGCANAGAGACTAGAGCACTGATGATCAATAAGAAGCTT 385
Oy	182 ataactttgaccaaatatttgccgtagagaccggaaatttgttgcgttcacaagccaagaat 241 
Db	386 CTTCCTCGAGAGATAGATTGATTATGATTGATTCGAGAAATTAATTAATGCCAANAAGAT 445
Oy	242 tlaatgcagcgctgagtcctgtlaaaggatgatlgaaacttlttggtccta 294 
Db	446 CTCAGCGTCGAACCAATGCCGAGATATTGANTGTGAACTTCTAAGATGAT 498
RESULT 5	CNS00EAX 1101 bp DNA GSS 04-JUN-1999
CNS00EAX/C	Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION	BACR2BD24 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL068861
VERSION	AL068861.1 GI:4949105
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster Eukaryota; Neoptera; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 1101)
REFERENCE	Genoscope.
AUTHORS	Direct Submission
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL	BP 191 9106 EVRY cedex - FRANCE (E-mail : seqret@genoscope.cns.fr - Web : <a href="http://www.genoscope.cns.fr">www.genoscope.cns.fr</a> )
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osonegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,



QY 158 ccgctatagtgattatgcaggtataacttgacaaaalltggcgtagaaaccgaat 217  
LOCUS AW755195 388 bp mRNA EST 01-MAY-2000  
DEFINITION PC12B08 Pine Triplex pollen cone library Pinus taeda cDNA clone  
ACCESSION AW755195  
VERSION AW755195.1 GI:7676915  
KEYWORDS EST  
SOURCE Pinus taeda  
ORGANISM Lobloily pine.  
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.  
1 (bases 1 to 388)  
AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.  
TITLE The Pine Gene Discovery Project  
JOURNAL Unpublished (1999)  
COMMENT Contact: Ross Whetten  
Forest Biotechnology Group  
North Carolina State University  
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh  
NC, 27695-8008  
Tel: 919-515-7800  
Fax: 919-515-7801  
Email: rosswhetten@unity.ncsu.edu  
Seq primer: 5' lambda Triplex2 Sequencing Primer.  
FEATURES  
source Location/Qualifiers  
1..388  
/organism="Pinus taeda"  
/db\_xref="taxon:3352"  
/clone="PC12B08"  
/clone\_lib="Pine Triplex pollen cone library"  
/dev\_stage="Immature"  
/note="Organ: pollen cone; Vector: Lambda Triplex; Site:1:  
SfiI (A); Site:2: SfiI (B); Immature pollen cones were  
collected in the early spring, frozen and used for mRNA  
isolation. The SMART-PCR method (Clontech) was used to  
prepare a library from 1 ug total RNA, using the Lambda  
Triplex vector. Plasmid subclones in pTriplex were  
recovered by cre-lox excision in E. coli strain BM25.8 and  
sequenced from the 5' end."

BASE COUNT 80 a 104 c 88 g 102 t 14 others  
ORIGIN

Query Match 6.7%; Score 36.6; DB 94; Length 388;  
Best Local Similarity 56.4%; Pred. No. 2;  
Matches 57; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 167 gtaattatgcaggtataacttgacaaaalltggcgtagaaaccgaattgtgtgt 226  
DB 74 GTTTTNNNNNNNTTATACATGNNNAATTGAGTATCAACGGAACGTGGTGG 133  
QY 227 cagagcccaagaatttaatgcagggtagtcctgtgtaaa 267  
DB 134 GTGGCTCGATGATNTTATGCTGTGGTGGCCATGCACAAA 174  
RESULT 9

AQ256569  
LOCUS AQ256569 587 bp DNA GSS 23-OCT-1998  
DEFINITION nbx0016B19r CUGI Rice BAC library Oryza sativa genomic clone  
nbx0016B19r, DNA sequence.  
ACCESSION AQ256569  
VERSION AQ256569.1 GI:3781051  
KEYWORDS GSS.  
SOURCE Oryza sativa.  
ORGANISM Oryza sativa  
REFERENCE Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.  
1 (bases 1 to 587)  
AUTHORS Wing,R.A. and Dean,R.A.  
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome  
JOURNAL Unpublished (1998)  
COMMENT Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seq primer: GGAAACGCTATGACCATG  
Class: BAC ends  
High quality sequence stop: 413.  
FEATURES  
source Location/Qualifiers  
1..587  
/organism="Oryza sativa"  
/strain="Japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:4530"  
/clone="nbx0016B19r"  
/clone\_lib="CUGI Rice BAC Library"  
/tissue\_type="Leaf"  
/lab\_host="E. coli DH10B"  
/note="Vector: pBlotBAC11; Site:1: HindIII; Site:2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from Oryza sativa, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 kb providing 10.9 haploid genome equivalents. The  
deep coverage allows the isolation a particular sequence  
with a probability of 99.9 %. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."

BASE COUNT 183 a 128 c 112 g 164 t  
ORIGIN

Query Match 6.7%; Score 36.6; DB 151; Length 587;  
Best Local Similarity 56.1%; Pred. No. 2,3;  
Matches 69; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 309 taacttcatacaccattatgcagcaaaattaggaattgctaaagct 368  
DB 460 TCAATTAATTGTTATGCAATTATGCCAAGACTAATTAGTATTTCTTAGCACACAC 519  
QY 369 agatgttaccagccgtaatgcacatactcaacaacgcgacaacacgacctag 428  
DB 520 AAACGCAATCAATCGTCAATCGACCCACATGACACGGAACCCCTGATCACTAC 579  
QY 429 agg 431  
DB 580 ACG 582  
RESULT 9







QY 236 aagaattlaatgcaagcgtagctcctgtaaaaggtgatgtagaaglcctlllgtgcttat 294  
||| | || | |  
Db 150 GAGAACTAAAGCGCTGGAACCAAGGTACTATTCACCTGACACATGAATCCTTAAAGAGTGGAT 208  
||| | || | |

Search completed: March 5, 2001, 03:19:25  
Job time: 19937 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 6, 2001, 14:36:12 ; Search time 26.49 Seconds  
(without alignments)  
232.347 Million cell updates/sec

Title: US-09-164-714-7  
Perfect score: 912  
Sequence: 1 MKTLKTLAVASASLLAMSA.....SYNYLSEDAANISGLHLP 180

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues  
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: A\_Geneseq\_36:\*  
2: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	912	100.0	180	21	Y84612
2	209	22.9	40	21	Y84611
3	152	16.7	31	21	Y56863
4	115	12.6	188	13	R26415
5	112.5	12.3	361	11	R05999
6	112.5	12.3	361	16	R66877
7	112	12.3	359	16	R66294
8	112	12.3	362	10	P90098
9	108	11.8	188	13	R25674
10	108	11.8	188	16	R70744
11	108	11.8	188	18	W18380
12	106.5	11.7	162	17	R96207

13	106.5	11.7	338	16	R85450
14	106.5	11.7	568	17	R96210
15	100	11.0	288	18	R7720
16	100	11.0	1287	16	R79944
17	97	10.6	211	18	M55324
18	97	10.6	211	18	M20755
19	95.5	10.5	1290	19	M98269
20	95	10.4	1288	18	M55547
21	95	10.4	1288	18	M55685
22	94.5	10.4	390	18	M21678
23	93.5	10.3	287	18	M53462
24	91	10.0	1296	14	R41198
25	89.5	9.8	767	19	M46272
26	89.5	9.8	1267	11	R04232
27	88.5	9.7	174	17	M04891
28	87.5	9.6	443	20	M94460
29	86.5	9.5	363	16	R66878
30	85.5	9.4	711	19	M73025
31	85.5	9.4	758	19	M46270
32	85	9.3	621	18	M55645
33	85	9.3	621	20	M17209
34	85	9.3	637	18	M55739
35	85	9.3	637	20	M17208
36	85	9.3	691	19	M73032
37	85	9.3	691	19	M73032
38	84.5	9.3	255	18	M55477
39	84.5	9.3	255	18	M55201
40	84.5	9.3	255	20	M17205
41	84	9.2	174	17	M04893
42	83	9.1	759	19	M46271
43	83	9.1	759	19	M46273
44	83	9.1	759	21	M52396
45	83	9.1	759	21	M52136

ALIGNMENTS

RESULT 1	
ID	Y84612
XX	Y84612 standard: Protein: 180 AA.
AC	Y84612;
XX	
DT	25-JUL-2000 (first entry)
XX	
DE	The outer membrane protein 21 of strain ATCC49143.
XX	
KW	Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection; otitis media; respiratory infection; sinusitis; pneumonia; immunisation.
XX	
OS	Moraxella catarrhalis.
XX	
FH	Key
FT	Misc-difference 71
FT	Location/Qualifiers
XX	/note="Ala encoded by CCC"
PN	WO200018910-A1.
XX	
PD	06-APR-2000.
XX	
PF	01-OCT-1999; 99WO-US22918.
XX	
PR	01-OCT-1998; 98US-0164714.
XX	
PA	(ANTE-) ANTEX BIOLOGICS INC.
XX	
PI	Tucker K, Tillmann UF;
XX	
DR	WPI: 2000-293149/25.
XX	
DR	N-PSDB: A12591.
XX	
PT	Isolated outer membrane protein from a Moraxella catarrhalis strain

PT used for diagnosis treatment and prevention of disease caused by M.  
PT catarrhalis e.g. pneumonitis, otitis media and respiratory infections -  
XX  
XX  
PS Claim 2; Fig 4; 108pp; English.  
XX  
CC The present sequence represents an outer membrane protein 21 (OMP21)  
CC of Moraxella catarrhalis strain ATCC91943. The OMP21 protein has an  
CC apparent molecular weight of 16-20 kd as determined by sodium  
CC dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). OMP21,  
CC its nucleic acids and antibodies can be used in prophylactic and  
CC therapeutic compositions for treating a M. catarrhalis bacterial  
CC infection, otitis media, respiratory infections, sinusitis and  
CC pneumonia. They are useful as reagents for the clinical or medical  
CC diagnosis of M. catarrhalis infections and for scientific research on  
CC the properties of pathogenicity, virulence and infectivity of  
CC M. catarrhalis and host defence mechanisms. The antibodies, particularly  
CC those that are cytotoxic may be used in passive immunisation to prevent  
CC or attenuate M. catarrhalis infections of animals e.g. humans.  
CC

Sequence 180 AA;

Query Match	100.0%	Score 912;	DB 21;	Length 180;
Best Local Similarity	100.0%	Pred. No. 2.0e-82;		
Matches 180; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

```
Oy      1 MktLTLLAASASSLLAMSAANAISYGNADAPQPYVGAKIGVDAKQINGKNTATGYAC   60
        |||||
        |||||
        |||||
Dbb     1 mktlktllavsassllamsanaaisygnadapqpyvakiqvdaqnknkntayiyaq   60
```

QY 61 YNFDQNGVEAEFVGSDAKEFNAGVSPVKGDVKSFGAYGTYRYNFINTPFYAKGKLGIAK 120

Db 61 ynfddnfgveaefvgsdakēfnagvspvkdvksfgaygtlyrnfintpfyakgk lgiak 120

Qy	121	TKVDVTSRNA	TTYSNKS	SDKTS	LAGGV	GVGFK	PLANV	GEASYN	YLSE	DANAI	SLGA	HLAF	180
Db	121	tkvdvt	srnat	tyns	ksd	tsl	agvg	vgv	fkp	lan	qv	ee	syn
	180	yl	se	dan	ai	sl	ga	hl	af				

RESULT	2
Y84611	
ID	Y84611 standard; peptide; 40 AA.

AA	
AC	Y84611;
XX	
DT	25-JUL-2000 (first entry)

DE N-terminal peptide of outer membrane protein 21 of strain ATCC49143.

KM Outer membrane protein 21; OMP21; strain ATCC49143; bacterial infection;  
KM otitis media; respiratory infection; sinusitis; pneumonia; immunisation.

05 Moraxella catarrhalis.

PN WO200018910-A1.

PD 06-APR-2000.

01-OCT-1999; 99WO-US22918.

PR 01-OCT-1998; 98US-0164714.

PA (ANTE-) ANTTEX BIOLOGICS INC.

PI Tucker K, Tillymann UF;

DR WPI; 2000-293149/25.

PT Isolated outer membrane protein from a *Moraxella catarrhalis* strain

PT catarrhalis e.g. pneumonia, otitis media and respiratory infections -

PS Claim 2; Page 63; 108pp; English.

XX The present sequence represents the N-terminal peptide of a mature  
CC outer membrane protein 21 (OMP21) of *Moraxella catarrhalis* strain  
CC ATCC49143. The specification describes an OMP21 from a *Moraxella*  
CC *catarrhalis* strain, which has an apparent molecular weight of 16-20 kD  
CC as determined by sodium dodecylsulfate polyacrylamide gel electrophoresis  
CC (SDS-PAGE). OMP21, its nucleic acids and antibodies can be used in  
CC prophylactic and therapeutic compositions for treating a *M. catarrhalis*  
CC bacterial infection, otitis media, respiratory infections, sinusitis  
CC and pneumonia. They are useful as reagents for the clinical or medical  
CC diagnosis of *M. catarrhalis* infections and for scientific research on  
CC the properties of pathogenicity, virulence and infectivity of  
CC *M. catarrhalis* and host defence mechanisms. The antibodies, particularly  
CC those that are cytotoxic may be used in passive immunisation to prevent  
CC or attenuate *M. catarrhalis* infections of animals e.g. humans.  
XX

5Q Sequence 40 AA

```
Query Match      22.9%; Score 209; DB 21; Length 40;
Best Local Similarity 100.0%; Pred. No. 4.4e-14;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy      23 AISTGNSADQPYVGAKIGOVDAKQINGKNTAYGIYAGYN 62
          |||||
Db      1 aistgnsadaqpyvgakigqvdaqngkntaygiyagyn 40

```

RESULT	3
156863	

XX	
XX	Y56863;
AC	
XX	10-APR-2000 (first entry)
DT	

DE B. catarrhalis 20 kD antigen N-terminal fragment

KW Brahanmella catarrhalis; Moraxella catarrhalis; antigen; vaccine immunogenic; respiratory tract infection; otitis media.

OS Branhamella catarrhalis

PN W09958563-A2.

PD 18-NOV-1999

PF 11-MAY-1999; 99WO-GB01473.

PR 11-MAY-1998; 98GB-0010084.

PA (CORT-) CORTECS UK LTD.

PI Cripps AW, Kyd J,

DR WPI; 2000-116286/10.

PT Novel antigens of *Branhamella catarrhalis* used for diagnosis, detection

XX XX

CC The invention relates to novel *Branhamella catarrhalis* (Moraxella  
CC *catarrhalis*) antigens having an apparent molecular weight of about 14-17  
CC kDa (as determined by SDS-PAGE). The antigens can be used to prepare  
CC vaccines and immunogenic compositions for the treatment and prophylaxis  
CC of *Branhamella catarrhalis* infections, respiratory tract infections, and  
CC otitis media. Antibodies against the antigens can be used for diagnosis  
CC and purification of the antigens. Sequences Y5683-871 represent peptidic  
CC peptide fragments from the *B. catarrhalis* antigens of the invention.

5Q	Sequence	31 AA;
----	----------	--------

Query Match	16.7%	Score 152	DB 21	Length 31
Best Local Similarity	96.8%	Pred. No. 1.3e-08		
Matches 30	Conservative 0	Mismatches 1	Indels 0	Gaps 0
QY	23	ALSYGNSADAPQYVGAKTGQVDAKINGKNT	53	
Db	1	alsygnssadapqyvgaklgyvdaklknkt	31	

RESULT	4
R26415	
ID	R26415 standard; Protein; 188 AA.

DT 17-MAR-1993 (first entry)  
XX  
XX pagC/AP fusion protein.  
DE  
XX  
XX Salmonella; pagC; alkaline phosphatase; AP; transposon; TnpHoA;  
KW Salm119; phoP; regulatory region regulated gene; virulence.  
XX  
XX Salmonella typhimurium.  
DS

FT	Key	Location/Qualifiers
FT	Peptide	1..23
FT	Protein	/label= Signal-peptide 24..188
FT	Region	/label= Mature-protein 1..33
FT	Region	/label= pagC 34..188
FT	Region	/label= AP

PN W09217785-A.

PD 15-OCT-1992.

PF 30-MAR-1992; 92WO-US02591.

PR 29-MAR-1991; 91US-0678409.

PA (GEHO ) GEN HOSPITAL CORP.

PI Miller SI;

DR WPI; 1992-366389/44.

XX

PT Detection of Salmonella using monoclonal antibodies - against Pho  
PT P regulated gene product useful for diagnosing e.g. typhoid  
PT fever, enteric fever(s) acute gastroenteritis etc.

PS Disclosure; Fig 2; 39pp; English.

CC This sentence represents a fusion between the *Salmonella* pagC protein  
CC and the alkaline phosphatase (AP) protein. The AP gene was inserted  
CC into the pagC gene on the transposon TnpH $\alpha$ . Insertion of this  
CC transposon into pagC mutant strain CS119 caused interruption of an  
CC 18kD protein to form the new protein species of approx. 45 kD given.  
CC pagC is a phop regulatory region regulated gene and a mutation at the  
CC pagC locus confers a virulence defect on *Salmonella* typhimurium.

Sequence 188 AA:

```
Query Match      12.6%; Score 115; DB 13; Length 188;
Best Local Similarity 27.1%; Pred. No. 0.00058;
Matches 55; Conservative 28; Mismatches 82; Indels 38; Gaps 11,
```

```

Db      1 mknlllslvtltslvvnvngadnatsvgyapyaqskv--gdfrklirgynvnyk-----f 54

QY      61 YNFDFNGFVEAEFWGSDA-----KEFNAGVSPV-----KGDVKSFGAGCT-----YRY 103
        | : : : : : | : : : : : | : : : : : |
Db      55 yedds-----pvsfslslsylyvdrtqasgsvepegihydkrevk-----ygslnmgpyrll 106

QY      104 NFINTPFPAKAGKLGIAK-TKVDVTSRNATYTSNK--SDKTSISLAGVGVGKRPPLANNVEA 150
        | : : : : : | : : : : : | : : : : : | : : : : : |
Db      107 s-dnslsyalagvgtvkafckehstqdgdsfsnkissrctgfawagavgmpleniavdv 165

QY      161 SY----NYLSEDAANISLGAHLAF 180
        | : : : : : | : : : : : |

Db      166 gyegsnisslcklingfnvgyvgyrf 188

```

RESULT	5
R05999	
ID	R05999 standard; protein; 361 AA

AC R05999;

DT 04-DEC-1990 (first entry)

P2 gene product of Haemophilus strain 1H.

P2 gene; strain 1H; influenza; vaccine; ds.

OS Haemophilus influenzae.

PN EP378929-A.

PD 25-JUL-1990.

PF 22-DEC-1989; 89EP-0313573.

PR 01-FEB-1989; 89GB-0002178.

XX

XX

XX

DR N-PSDB; Q05372.

P1<sup>1</sup> Gene coding for

PT caused by the organism.

PS Claim 1; Page 8; 15pp; English.

CC Peptide may be used as a vaccine to the disease caused by H1  
CC type b, as carrier for conjugation to oligosaccharide derived  
CC from Haemophilus. Protein may be used with other haptens as  
CC T-cell dependant antigen and carrier.

Sequence 361 AA;

Query Match	12.3%	Score 112.5	DB 11	Length 361
Best Local Similarity	28.1%	Pred. NO. 0.0024		
Matches 52; Conservative	25;	Mismatches 65;	Indels 43;	Gaps 11;

[illegible][illegible][illegible][illegible]



QY 99 -GTYRYNFTNPFYAKGLGIKTKVD---VTSRNATTVSNKSDKTSIAGC---VGVGFK 151  
 Db 117 kgsye---vldgldvygkagvalvrsdykfyledangtrchkhkgrhcarasglfavgaeys 173  
 QY 152 PLANVGEASYNVLS 166  
 Db 174 vlpevlavrlayqwt 188

## RESULT 8

P90098  
 ID P90098 standard; protein; 362 AA.

AC P90098;

DT 01-NOV-1989 (first entry)

XX P2 antigen of Haemophilus influenzae.

XX Haemophilus influenzae P2 antigen; vaccines; pathogenic  
 KM type b strains.

XX Haemophilus influenzae type b.

XX EP320289-A.

XX 14-JUN-1989.

PF 09-DEC-1988; 88EP-0311691.

XX 10-DEC-1987; 87US-0249482.

XX (TEXA ) UNIV OF TEXAS SYST.

XX Hansen EJ;

XX WPI; 1989-174562/24.

XX DNA encoding Haemophilus influenzae P2 antigen

PT - used for producing immunogenic compns. as vaccines  
 PT and in diagnostic hybridisation assays.

XX Disclosure; fig 6; 20pp; English.

XX P2 antigen of Haemophilus influenzae (see corresp. N90033).

CC Isolated from type b strains, the P2 antigens are used as vaccines  
 CC and to detect P2-specific RNA or DNA. Segments used as antigens

CC Include amino acids 21-361, 175-197, 260-275, 296-311, 33-353,  
 CC or their functional equivalents.

XX Sequence 362 AA;

Query Match 12.3%; Score 112; DB 10; Length 362;  
 Best Local Similarity 28.0%; Pred. No. 0.0027;

Matches 52; Conservative 25; Mismatches 65; Indels 44; Gaps 11;

QY 5 KTLTAAVSASSLLAMSAANAISYNSADAPYGAIRIGOV-----DAKQING----- 50  
 Db 3 KTLAALIVGAFASAAANAAYVYNEG-tnevlgrisltaegsnstvcdnqkqhgahln 61  
 QY 51 KNTAVGIVGYNEFDQNF---GVEAEFV-----GSDAKEFNAGVSPVKGDKVSGFAYGT 100  
 Db 62 qgsrflfkakthnqgdgfyagyltrfvtckasengsd--n-----gdls-----k 106  
 QY 101 YRYNFTNPFYAKGLGIKTKVD-VTSRNATTVS--NKSDKTSLAGG-VGVGFKPLANY 156  
 Db 107 yayvltgkafgevkigrakltiadgiltseadkeygvlmsdyiptsgntvgytlfkglidgi 166  
 QY 157 GVEASY 162  
 Db 167 vlgany 172

RESULT 9  
 R25674  
 ID R25674 standard; Protein; 188 AA.

XX R25674;

DT 13-JAN-1993 (first entry)

XX pagc.

XX Salmonella positive regulon; phop; vaccination.

XX Salmonella typhimurium.

XX WO9211361-A.

XX 09-JUL-1992.

PF 18-DEC-1991; 91WO-US09604.

XX 18-DEC-1990; 90US-0629602.

XX (GEHO ) GEN HOSPITAL CORP.

XX (HARD ) HARVARD COLLEGE.

XX Mekalanos JJ, Miller ST;

XX WPI; 1992-250084/30.

XX N-PSDB; Q26694.

XX New anti-salmonella vaccine contg. live attenuated mutants -  
 PT useful to protect against typhoid fever and related diseases, and  
 PT for detecting salmonella

XX Disclosure; Page 49-51; 57pp; English.

XX pagc is a Salmonella positive regulon (phop) regulated gene. A

CC mutation at pagc confers a virulence defect. Within the scope of

CC the invention, this gene was mutated, such that it included a

CC virulence attenuating mutation. This mutation does not result in

CC the constitutive expression of a gene under the control of the phop

CC regulatory region. The vector sequence encoding this sequence may

CC be used to transform a cell and then the pagc gene product produced.

CC The mutated pagc gene can be used to produce a vaccine comprising a

CC Salmonella cell which has been attenuated by the above mutation.

CC These vaccines can be administered orally thus avoiding resistance

CC by patients to vaccination by injection.

XX Sequence 188 AA;

Query Match 11.8%; Score 108; DB 13; Length 188;

Best Local Similarity 24.2%; Pred. No. 0.0028;

Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

QY 1 MKTLKTLTAAVSASSLLAMSAANAISYNSADAPYGAIRIGOVDAKQINGKNTAYGIYAG 60  
 Db 1 MKNILLSTVILTSVLYVNVNAGADTNAFSVYARYAGSKV--qdfknlitgvnnky----- 53  
 QY 61 YNFDQNFGEVGAEPFVGSADKEFNAGVSPVKGDKVSGFCA-----YGT----- 100  
 Db 54 -----ryedaspvflisslslylgydrqagsvvepiglihndkfevkygslmvgp 102  
 QY 101 -YRYNFTNPFYAKGLGIKTKVD-VTSRNATTVS--SDKTSIAGGCVGVGFKPLANY 156  
 Db 103 ayrls-dnfslyalagvgtvkafkeshstqgdsfsnkissrktgtfawgvgvgnmpleni 161  
 QY 157 GVEASY---NYLSEDAANAISLGAHLAF 180  
 Db 162 vvdvgyegsnisscktingfnvgvyrtf 188

SEQUENCE	DESCRIPTION
10	RESULT
11	R70744
12	R70744 standard; Protein; 188 AA.
13	R70744;
14	14-AUG-1995 (first entry)
15	PagC protein.
16	Salmonella; vaccine; attenuation; pagC; Phop regulon.
17	Salmonella typhimurium.
18	W09502048-A.
19	19-JAN-1995.
20	07-JUL-1994; 94MO-US07658.
21	09-JUL-1993; 93US-0090526.
22	06-JUL-1994; 94US-0271354.
23	(GEHO ) GEN HOSPITAL CORP.
24	(HARD ) HARVARD COLLEGE.
25	Mekalanos JJ, Miller SI;
26	WPI: 1995-066894/09.
27	N-PSDB: C85099.
28	Salmonella vaccines contg. cells attenuated by mutation in a Phop
29	regulon - are highly attenuated and very safe; partic. for oral
30	admin.; mutation may opt. also be in aromatic amino acid
31	synthesis gene.
32	Disclosure: Page 114-115; 182pp; English.
33	The phop regulatory region activated gene (pagC) pagC of wild-type
34	Salmonella typhimurium was cloned and sequenced (085099). The pagC
35	protein (R70744) is essential for survival within macrophages and
36	for virulence; mutation of pagC resulted in attenuation by at least
37	1,000-fold.
38	Sequence 188 AA;

```

Query Match      11.8%: Score 108; DB 16; Length 188;
Best Local Similarity 24.2%: Pred. No. 0.0028;
Matches 50; Conservative 29; Nonmatches 82; Indels 46; Gaps 8;

QY      1 MKTLTLLAVASSILLASANAATSYGASADAPPYGVAKIQGVDAKQINGKRTATAGIAG 60
          ||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1 mknl|stclvtlts|lvvnvagaqadlnafsvyaraqskv--qdfknirgvnvkY----- 53

QY      61 YNFDPNGFVEAEFVGSDAKEFNAGSPVPKGVKSPGCA-----YGF----- 100
          :||::||::||::||::||::||::||::||::||::||::||::||
Db      54 -----ryeddpvsf|ss|sy|ygdraqsgvspeepg|nyhdkfeykyslmvyp 102

QY      101 -YRNFINTPEPYANGKLGIAK-TKVDYTSRNTATYSNK--SDKTSLAGVGGVGFPLANV 156
          ||::||::||::||::||::||::||::||::||::||::||::||::||
Db      103 ayr|s-dnfs|yalagvgtkaftebhs|tdgddgs|snk|ssrktg|fawgagvqm|pleni 161

QY      157 GVEASY---N|LSEDA|NALS|GAHLAF 180
          |::||::||::||::||::||::||::||::||::||::||::||::||
Db      162 vvdvgyegsn|ss|stekingfnvgvyrf 188

RESULT  11
W18380
ID      W18380 standard; Protein; 188 AA.
FX

```

AC	WI8380,
XX	
DT	27-AUG-1997 (first entry)
XX	
XX	S. typhimurium pagC gene'expression product.
DE	
XX	
KW	Salmonella. vaccine; attenuation; diagnosis; pagC gene; Phop;
KW	typhoid fever; virulence.
XX	
OS	Salmonella typhimurium ATCC 10428.
XX	
FH	Key
FT	Location/Qualifiers
FT	1.23
FT	/label= Sig-peptide
XX	
PN	WO9720033-A1.
XX	
PD	05-JUN-1997.
XX	
PF	27-NOV-1996; 96WO-US19190.
XX	
PR	01-DEC-1995; 95US-0565861.
XX	
PA	(GEO ) GEN HOSPITAL CORP.
PA	(HARD ) HARVARD COLLEGE.
XX	
PI	Mekalanos JJ, Miller SI;
XX	
DR	WPI; 1997-310590/28.
DR	N-PSDB; T70212.
XX	
PT	Attenuated Salmonella strains containing PhoQ or Phop mutation(s) -
PT	used in vaccines, for protection against Salmonella infection and,
XX	e.g. prevention and treatment of typhoid fever
XX	
PS	Disclosure; Fig 3; 176pp; English.
XX	
CC	The 186-amino acid expression product (WI8380) of the Salmonella
CC	typhimurium pagC gene (T70212) is a membrane protein essential for
CC	S. typhimurium survival within macrophages and for virulence.
CC	It shows sequence similarity to the phage lambda protein lam and
CC	to the ail gene product of Yersinia enterocolitica, Salmonella
CC	pad mutant strains are attenuated for virulence and can be used as
CC	live Salmonella vaccines.
XX	
Sequence	188 AA;

```

Query Match      11.8%; Score 108; DB 18; Length 188;
Best Local Similarity 24.2%; Pred. NO. 0.0028;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

OY      1 MKTLTLLAVSAASSILLANSANAAISVGNSSADQPIYGAKICGVDAKQINGKRITATGIAG 60
        |||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       1 mnullstlvtlsylvnvagadnatsfvgayraqskv--qdklnlrnyvk----- 53

OY      61 YNFDONFCGEAEFFVSADKEFNAGSVPAVKGDVKSFGA-----YGT----- 100
        :||::||::||::||::||::||::||::||::||::||::||::||::||
Db       54 -----ryeddspsvfisslslylgdrqaasveepgihyhdkfevkrgslmvgp 102

OY      101 -YRNFINTPEYFAKGLGIAT-TKDVTISRNAATTYSNK--SPDKTSLAGVGCGFPLANV 156
        ||::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       103 ayrll-dnfslayalagyvtlkatfcfhetsqdqdsfsnktsrktgfawgavgvmp]cni 161

OY      157 GVEASY---NYLSEDNAISLGAIHAF 180
        !::||::||::||::||::||::||::||::||::||::||::||::||::||
Db       162 vdvdygegsnlstckngfmvgvyrf 188

RESULT 12
R96207
ID R96207 standard; Protein; 162 AA.
XX
```

```

AC R96207;
XX
XX 24-AUG-1996 (first entry)
DE Attachment-Invasion-locus protein.
XX
XX Attachment-Invasion-locus protein; drug delivery;
KM gastrointestinal membrane; transcytosis; bioavailability;
KM fusion protein; enterocyte; Peyer's patch M-cell.
XX
OS Versinia enterocolitica strain 8081c.
XX
XX Key Location/Qualifiers
FH Peptide 19..34
FT /note= "Peptide from extracellular loop-1"
FT 58..76
FT /note= "Peptide from extracellular loop-2"
FT 102..119
FT /note= "Peptide from extracellular loop-3"
FT 145..152
FT /note= "Peptide from extracellular loop-4"
XX
XX MO9613250-A1.
XX
XX 09-MAY-1996.
XX
XX 20-OCT-1995; 95MO-US13749.
XX
XX 27-OCT-1994; 94US-0331393.
XX
XX (AMGE-) AMGEN INC.
XX
XX Habberfield AD, Jensen-Pippo K;
XX
XX WPI: 1996-251447/25.
XX
XX N-PSDB: T27536.
XX
XX
XX Therapeutic delivery system utilising bacterial invasion protein - is
PT not readily degraded in the gut, enhances systemic bio-availability
PT of therapeutic agents
XX
XX Example 1; Fig 2; 110pp; English.
XX
XX The sequence represents a bacterial attachment-invasion-locus (AIL)
CC protein, which may be complexed with a therapeutic agent to
CC transport the agent across the gastrointestinal membrane barrier by
CC transcytosis to increase bioavailability 5- to 100-fold. The agent
CC and bacterial protein may be linked via a degradable peptide
CC sequence as a fusion protein. The receptor binding region of the
CC AIL protein involves all or some of the regions from the 4
CC extracellular loops. This region retains the binding affinity of
CC the protein, and may be used alone or as part of a fusion protein
CC for drug delivery. The protein may be fused with maltose binding
CC protein (R96208) to form a fusion protein (R96209) which may be
CC purified easily by amylose affinity chromatography. The delivery
CC system allows improved transport across enterocytes and Peyer's
CC patch M-cells. The system is not prone to degradation in the gut or
CC early release of biologically active material, and eliminates the
CC need for parenteral administration.
XX
XX
XX Sequence 162 AA:

```

```

Query Match 11.7%; Score 106.5; DB 17; Length 162;
Best Local Similarity 30.1%; Pred. No. 0.0033;
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

```

```

OY 10 VSASSLJAMSANAISY---GNSADAPYVGAKTGVDAKQNGKNTATYAGVNFQ 65
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 5 vyaaesessisigyagshvkenytlndp-----kgfllky-----ryeld 46
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 66 NFGVEAEFVGS-DAKEFNAGVSPV-KGDVKSFGAYGTYRYNF-INPFFYAKGLIAKTK 122
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

```

Db 47 nwgvgisfaytlbggydfygsnkgfhgdwdyvs--tmgpsfirneyvalygl1gaahg 104
OY 123 VDVTSRNATTY---SNKSDSTSLAGVGGEKPLANYGVEASYNYLSEDANNAIS---LGAH 177
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 105 v-----kasvfdeslsastksmaysagvgfnlpnlvldasyskldskvgtm1gag 159
OY 178 LAF 180
|
Db 160 yrf 162

```

```

Query Match 11.7%; Score 106.5; DB 16; Length 338;
Best Local Similarity 25.6%; Pred. No. 0.0086;
Matches 50; Conservative 18; Mismatches 64; Indels 63; Gaps 11;

```

```

OY 35 YVGAKTQ---VDAKQING-----KNT-AYGIYAGYNF--DONFGVEAE 72
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 8 yagvkagqgsfhdginnngakidsldltlygyryntltfyfgyqllngdnfglaae 67
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
OY 73 FVGSID-----AKEFNAGVS-PVKGDVKSFGAYGTYRYNF-INPFFYAKGLK 117
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```







**THIS PAGE BLANK (USPTO)**



OY		99	-GTRRYNINIRPFAKKGGLGIATKKVD---VTSNATTYSNKSPKSLAGC--- <td>151</td>	151
			:   :   :   :   :   :   :   :	
Dd		117	KGSYE---VLGDLDVYGKAGVALYRSDEKYEDANGTDRHKKGHTTKARASGLFAYGAETA	173
			:   :   :   :   :   :   :   :	
OY		152	PLANVGVEASTNYLS	166
			:   :   :   :   :   :   :   :	
Ld		174	VLPBELAVRLELYQWLT	188

RESULT 2  
US-08-467-722A-2

```

1      Sequence 2, Application US/08467722A
2      Patent No. 6030626
3
4      GENERAL INFORMATION:
5      APPLICANT: Kolatlukudy, P. E.
6      TITLE OF INVENTION: Otitis Media Vaccine
7      NUMBER OF SEQUENCES: 6
8      CORRESPONDENCE ADDRESS:
9      ADDRESSEE: Calfee, Halter and Griswold
10     STREET: Suite 1800 800 Superior Avenue
11     CITY: Cleveland
12     STATE: Ohio
13     COUNTRY: U.S.A>
14     ZIP: 44114-2688
15
16     COMPUTER READABLE FORM:
17
18     MEDIUM TYPE: Floppy disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.30
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/467,722A
24
25     FILING DATE:
26
27     CLASSIFICATION: 424
28
29     ATTORNEY/AGENT INFORMATION:
30     NAME: Goltjick, Mary E.
31     REGISTRATION NUMBER: 34,829
32     REFERENCE/DOCKET NUMBER: 8277/00102
33
34     TELECOMMUNICATION INFORMATION:
35     TELEPHONE: (216) 622-8458
36     TELEFAX: (216) 241-0816
37
38     INFORMATION FOR SEQ ID NO: 2:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 359 amino acids
41     TYPE: amino acid
42     TOPOLOGY: linear
43
44     MOLECULE TYPE: protein
45     OS-08-467-722A-2

```

Query Match	12.4%;	Score 113;	DB 3;	Length 359;
Best Local Similarity	28.2%;	Pred. No. 0.00014;		
Matches	55;	Conservative	18;	Mismatches 80;
			Indels	42;
			Gaps	11;

QY	5	KTLLASASSILLAMSAANAALISYGSADAGYVCAKIGQ---	VPAKQING-----	50
Db	3	KTALALVVGGLAASVAQAAPQEWTF---	YAGVKAQSGSEHDGINNNGIKKGLSSSNY	58
QY	51	---KNT-AYGIYAGYNF--DQNFGEAEFEVGSQ---	AKEEFNACVSPKGVKSEFAY---	98
Db	59	GYRNTTYYGVFGGYQLINDNFGGLAAELGTDGFRAKTLREACKPLAKH--	TTHGAYLSL	116
QY	99	-GTYYRNFINTPYAKGKGLIAKTKVD--	VTSRNATYYSNKSDKTSLAGG--	151
Db	117	KGSYE---VLDGDLVGVKAGVALVRSQDYKREYEDANGTRDHRKGRHTRARASGLAEVAAEYA		173
QY	152	PLANVGEASINTYLS	166	
Db	174	VLPETIAVRLLEYOWLT	188	

RESULT 3  
US-08-096-181A-8

; Sequence 8, Application US/08096181A  
; Patent No. 6153406

Query Match	12.3%;	Score 112.5;	DB 3;	Length 361;
Best Local Similarity	28.1%;	Pred. No. 0.00016;		
Matches 52;	Conservative 25;	Mismatches 65;	Indels 43;	Gaps 11;

[illegible]

```

RESULT      4
PCT-US94-08326-8
; Sequence 8, Application PC/TUS94.08326
;
; GENERAL INFORMATION:
;
; APPLICANT: North American Vaccine, Inc.
; APPLICANT: 12103 Indian Creek Court
; APPLICANT: Beltsville, MD 20705
; APPLICANT: Pullen, Jeffrey K.
; APPLICANT: Soper, Thomas S.
; APPLICANT: Liang, Shu-Mei
;
; TITLE OF INVENTION: A Method For The High Level
;
; TITLE OF INVENTION: Expression,
; TITLE OF INVENTION: Purification And Refolding Of The Outer Membrane

```

```

; TITLE OF INVENTION: Protein
; STATE: Massachusetts
; NUMBER OF INVENTION: P2 From Haemophilus Influenzae Type b
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/08326
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,181
; FILING DATE: 23-JULY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REFERENCE/DOCKET NUMBER: 1438.001PC01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US94-08326-8

Query Match      12.3%; Score 112.5; DB 4; Length 361;
Best Local Similarity 28.1%; Pred. No. 0.00016;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY 5 KTLVASASSLLAMSAANAISTGNSADAPYGAIGOV-----DAKQING---K 51
    ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3 KTLALIVAFAFASANAAMVYNNES--TNVELGRLSIIAEQSNSTVDKQKHGLRQ 61
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 52 NNAVGIYAGNFDQNF---GVAEFV-----GSDAKFEFNAGVSPVKDVSFSGAYGY 101
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 62 GSRPHKATHNFGDGYAGYLETFRVTKASENGSD--NF-----GDITS-----KY 106
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 102 RYNEINTPFYAKGLGIATKYD-VTSRNATYS--NKSDKTSLAGG-VGVGFKPLANYG 157
    | : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 107 AVYTLGNKAFGEVKGIRAKTIADGITSADKREYGVNNNDYIPTSGNIVGYTFKGDGLV 166
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 158 YEAST 162
    : : | |
DB 167 LGANY 171

RESULT 5
US-08-286-767-3
; Sequence 3, Application US/08286767
; Patent No. 5733760
; GENERAL INFORMATION:
; APPLICANT: Lu, Yichen
; APPLICANT: Miller, Samuel I.
; TITLE OF INVENTION: SALMONELLA VECTORS ENCODING TRUNCATED
; TITLE OF INVENTION: Pacc FUSION PROTEIN, METHOD OF MAKING, AND USES THEREO
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: RONALD I. EISENSTEIN; DIKE, BRONSTEIN,
; ADDRESSEE: ROBERTS & CUSHMAN
; STREET: 130 Water Street
```

```

; CITY: Boston
; STATE: Massachusetts
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,767
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Resnick, David S.
; REGISTRATION NUMBER: 34235
; REFERENCE/DOCKET NUMBER: 44420
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 523-3400
; TELEFAX: (617) 523-6400
; TELEX: 200291 STRE UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 188 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-286-767-3

Query Match      11.8%; Score 108; DB 1; Length 188;
Best Local Similarity 24.2%; Pred. No. 0.0002;
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

QY 1 KTKLTLVASASSLLAMSAANAISTGNSADAPYGAIGOVADAKQINGKNTANGIYAG 60
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1 MKNIILSTLVITTSYLVVNAQADTNAFSVGYARYQSKV--QDFKNINGVNNKY----- 53
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 YNFDQFVGEAEFVSGDAKEFNAGVSPVKGDVSKFGA-----YGT----- 100
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 54 -----RYEDDSPVFTSSLSLYLGDROAGSVEPEGCIHYHDKFEVKIGSLMVG 102
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 101 -RYNEINTPFYAKGLGIATK--TKVDVTSRNATYSNK--SPKTSIAGGVGVGFKPLANY 156
    || | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 103 AYRLS-DNFSILVALAGVGVYKATFKHSIQDDGDSFNKISSRKTGFAMGAGQVMNLENI 161
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 157 YEAST---NYLSEDAANAISLGAHLAF 180
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 162 VVDVGEYSNISSTKINGFNVGVGYRF 188
    : : | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
PCT-US95-13749-5
; Sequence 5, Application PC/TUS9513749
; GENERAL INFORMATION:
; APPLICANT: Amgen Inc.
; TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
; TITLE OF INVENTION: BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Inc.
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320-1789
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/13749
```

;; FILING DATE:  
;; CLASSIFICATION:  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 568 amino acids  
;; TYPE: amino acid  
;; STRANDEDNESS: unknown  
;; TOPOLOGY: unknown  
;; MOLECULE TYPE: protein  
PCT-US95-13749-5

Query Match 11.7%; Score 106.5; DB 4; Length 568;  
Best Local Similarity 30.1%; Pred. No. 0.0014;  
Matches 55; Conservative 19; Mismatches 72; Indels 37; Gaps 10;

QY 10 VSASSILMSANNAISY-----GNSADAPYVGAKIGVDAKQINKNPAYGIYAGYDFQ 65  
Db 411 YAAASESSISIGYQSHKENGTYLDNDP-----KGFNLKY-----RYELDD 452  
QY 66 NFGVAFVGS-DAKEFNAGVSPV-KGVKSGFAGYTRYRNF-INPPYAKGLGIATKR 122  
Db 453 NMGVGSFAYTHQGYDFFGSNKFGHGDVYYSV--TWGSPFRINEYVSLYGLLGAHGK 510  
QY 123 VDVSRNATTY--SNKSDKSLAGSGVGFKPLANVGVEASYNILSEDAANIS--LGAH 177  
Db 511 V-----KASVEDESISAKSTMAVGAGVQFNPDPFVIDASYSKLDSIKVGTWMLGAG 565  
QY 178 LAF 180  
Db 566 YRF 568

RESULT 7  
US-08-210-394-1  
; Sequence 1, Application US/08210394  
; Patent No. 5770213  
; GENERAL INFORMATION:  
; APPLICANT: Zlotnick Dr., Gary W.  
; TITLE OF INVENTION: Purified No. 5770213typable Haemophilus  
; TITLE OF INVENTION: Influenzae p5 Protein as a Vaccine for No. 5770213typable  
; TITLE OF INVENTION: Haemophilus Influenzae Strain  
; NUMBER OF SEQUENCES: 1  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: American Cyanamid Company  
; STREET: One Cyanamid Plaza  
; CITY: Wayne  
; STATE: New Jersey  
; COUNTRY: US  
; ZIP: 07470-8426  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/210.394  
; FILING DATE: 07-MAR-1994  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Harrington, James J  
; REFERENCE/DOCKET NUMBER: 32,144  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201/831-3246  
; TELEFAX: 201/831-3305  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 338 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-210-394-1

Query Match 11.6%; Score 105.5; DB 1; Length 338;  
Best Local Similarity 25.6%; Pred. No. 0.00085;  
Matches 50; Conservative 17; Mismatches 65; Indels 63; Gaps 11;

QY 35 YVGAKIGQ---VDAKQING-----KNT-AVGIYAGYNF--DQNGVEAE 72  
Db 8 YAGVACGCGSFHDGINNGAIKEDSIDLTGYYRRNFTFYGVFGYQILNQDNFGLAE 67  
QY 73 FVGSD-----AKEFNAGVS-PVKGVKSGFAGYTRYRNF-INPPYAKGLG 117  
Db 68 -LGDYNGFRVYFRAEGTKAKHTNHGAHLSLKGSTYELDLDY-----GKAG 114  
QY 118 IAKTRVD--VTSRNATTYSNKSDKSLAG--VGVGFKPLANVGVEASY-----NY 164  
Db 115 VALVSDYKFFEAPRSTDDKXKGTARASGLFANVAGYAVLPFLAVLLEFVQQLTRVGKY 174  
QY 165 LSEDAANISLGAHLLA 179  
Db 175 RPODKNAPSINPNTA 189

RESULT 8  
US-08-486-715-5  
; Sequence 5, Application US/08486715  
; Patent No. 5674724  
; GENERAL INFORMATION:  
; APPLICANT: Miller, Samuel I.  
; TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P. C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; COMPUTER: IBM PS/2 Model 502 or 55SX  
; OPERATING SYSTEM: MS-DOS (Version 5.0)  
; SOFTWARE: WordPerfect (Version 5.1)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/486.715  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/090.526  
; FILING DATE: 09-JUL-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/629.602  
; FILING DATE: 18-DEC-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Clark, Paul T.  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 00786/192005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 542-5070  
; TELEFAX: (617) 542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 188 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-486-715-5

Query Match 11.3%; Score 103; DB 1; Length 188;  
Best Local Similarity 27.0%; Pred. No. 0.00071;



Db 55 YEDDSPVSPFSSLSLYLGD-RQASGSVEPEGIHYHDKFEVKKGSLMVGPAFLS-DNFSL 112  
Qy 111 YAKKGLGIAT-KYVDVTSRNATTTYSNK--SDKTSIAGGVGCFKPLANVGVEASY---NY 164  
Db 113 YALAGVGTVKATFKFESHSTQDDSFSSNKISSRKTGFMAGAGVQMNPLENIIVDVGEGSNI 172  
Qy 165 LSEDANAISLGAHLAF 180  
Db 173 SSTKINGFNVGVGRF 188

RESULT 11  
US-08-475-749-5  
Sequence 5, Application US/08475749  
Patent No. 6010901  
GENERAL INFORMATION:  
APPLICANT: Miller III, Samuel I.  
APPLICANT: Mekalanos, John L.  
TITLE OF INVENTION: SALMONELLA VIRULENCE GENES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson P. C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" diskette, 1.44 Mb  
COMPUTER: IBM PS/2 Model 502 or 555X  
OPERATING SYSTEM: MS-DOS (Version 5.0)  
SOFTWARE: WordPerfect (Version 5.1)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08-475, 749  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/090, 526  
FILING DATE: 09-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/629, 602  
FILING DATE: 18-DEC-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Clark, Paul T.  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 00786/192004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 542-5070  
TELEFAX: (617) 542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 188 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-475-749-5

Query Match 11.3%, Score 103; DB 3; Length 188;  
Best Local Similarity 27.0%, Pred. No. 0.00071;  
Matches 53; Conservative 28; Mismatches 91; Indels 24; Gaps 10;  
Qy 1 MKLTLLAVASSSILLASANAISYGSNADAPYVGAIKIGVDAKOINGKMTATAGIYAG 60  
Db 1 MKNITLTLVITTSYLVVNVAAQADTNAPSVGIARYAQSKV--QDFKNIRGVAVKI-----R 54  
Qy 61 YNFDO--NFGVEAEFVGSDAKFEFNAGVSPV-----KGDVK--SFGAYGTYRYNINTPF 110  
Db 55 YEDDSPVSPFSSLSLYLGD-RQASGSVEPEGIHYHDKFEVKKGSLMVGPAFLS-DNFSL 112

Qy 111 YAKKGLGIAT-KYVDVTSRNATTTYSNK--SDKTSIAGGVGCFKPLANVGVEASY---NY 164  
Db 113 YALAGVGTVKATFKFESHSTQDDSFSSNKISSRKTGFMAGAGVQMNPLENIIVDVGEGSNI 172  
Qy 165 LSEDANAISLGAHLAF 180  
Db 173 SSTKINGFNVGVGRF 188

RESULT 12  
US-08-200-232-2  
Sequence 2, Application US/08200232  
Patent No. 5721349  
GENERAL INFORMATION:  
APPLICANT: Cover, Timothy L.  
APPLICANT: Blaser, Martin J.  
TITLE OF INVENTION: VACUOLATING TOXIN-DEFICIENT H. PYLORI  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NEEDLE & ROSENBERG P.C.  
STREET: 127 Peachtree Street, Suite 1200  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/200, 232  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Spratt, Gwendolyn D.  
REGISTRATION NUMBER: 36,016  
REFERENCE/DOCKET NUMBER: 2200, 023  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 404/688-0770  
TELEFAX: 404/688-9880  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1287 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-200-232-2

Query Match 11.0%, Score 100; DB 1; Length 1287;  
Best Local Similarity 24.8%, Pred. No. 0.022;  
Matches 53; Conservative 32; Mismatches 71; Indels 58; Gaps 12;  
Qy 9 AVSASSLLAMGSANAAISYGSNADAPYVGAIKIGQYDA-----KQIN----- 49  
Db 1025 AIGGAS-LNNGGNASL-YGTAGVADAYLN---GQVEALIVGSGSYGSSFFNNQANSLSNG 1079  
Qy 50 GKNTAYGIYAGYNPDQN-FGYEAE-FVGSDDAKFEFNAGVSPYKGDYKSF-----GA 97  
Db 1080 ANNTNFGVYSRIFANQHDFDEADGALGSDSSLSLFXSALRLDLQSYNYLAYSAATRAS 1139  
Qy 98 YGTYRYNFINTPFYAKGAGIATKTRVDVTSRNATTTYSNKSDKTSIAGGVGCFKPLANVG 157  
Db 1140 YG-YDFAFRNALVLRPSVGVSYNHLGSTNR-----SNSTKNVALSNSSSSOHLFNASAN 1194  
Qy 158 VEASVNY-----LSE-----DANAISL 174  
Db 1195 YEARYYYGDTSYFYWMNAGVLDGFANFGSSNAVSL 1228

RESULT 13







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:37:44 ; Search time 23.41 Seconds  
(without alignments)  
522.090 Million cell updates/sec

Title: US-09-164-714-7

Perfect score: 912  
Sequence: 1 MKTLKTLAVSASSLLAMSA.....SYNYLSEADANAISLGAHLAF 180

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

```

```
Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	123.5	13.5	178	2	F64124	opacity protein ho
2	123	13.5	178	2	A35123	ail protein precu
3	117	12.8	572	2	H81277	flagellin Cj139c
4	114.5	12.6	160	2	H82207	hypothetical prote
5	112.5	12.3	359	2	D64050	major outer membra
6	112.5	12.3	361	2	A30542	major outer membra
7	109.5	12.0	264	2	I54668	heat resistant aggl
8	108	11.8	188	2	A39185	virulence protein
9	108	11.8	365	1	MMBP2	outer membrane por
10	108	11.8	365	1	MMECNC	outer membrane por
11	103	11.3	371	2	S68072	major outer membra
12	103	11.3	371	2	S68069	major outer membra
13	102.5	11.2	213	2	A56152	major 25k outer mem
14	101	11.1	366	2	S68062	major outer membra
15	101	11.1	366	2	S68063	major outer membra
16	101	11.1	386	2	S68064	major outer membra
17	101	11.1	572	2	G81277	flagellin Cj138c
18	100	11.0	1287	2	B53739	vacuolating cytocto
19	100	11.0	1281	2	S44983	vacuolating cytocto
20	99.5	10.9	172	2	A39189	outer membrane pro
21	98.5	10.8	215	2	TC4087	accessory coloniza
22	97.5	10.7	1245	2	T49815	related to multifu
23	96	10.5	385	2	S68066	major outer membra
24	96	10.5	385	2	S68067	major outer membra
25	96	10.5	385	2	S68068	major outer membra
26	96	10.5	385	2	S68070	major outer membra
27	96	10.5	385	2	S68071	major outer membra
28	95.5	10.5	1290	2	G64630	vacuolating cytocto
29	95	10.4	193	2	T44441	hypothetical prote

30	95	10.4	21.1	2	S47347	outer membrane pro
31	95	10.4	57.5	2	S43310	flagellin A - Camp
32	95	10.4	57.5	2	I40615	flagellin protein
33	95	10.4	1288	2	E71884	vacuolating cytoce
34	92.5	10.1	22.1	2	A82331	hypochlorite prote
35	90	9.9	185	2	A43309	outer membrane pro
36	90	9.9	186	2	S70957	hypothetical prote
37	90	9.9	57.5	2	S41311	flagellin B - Camp
38	89.5	9.8	158.5	2	B69948	phage-related prot
39	89	9.8	17.1	2	I51173	outer membrane pro
40	89	9.8	359	2	S07298	outer membrane pro
41	88	9.6	36	1	WMECA	outer membrane pro
42	87	9.5	146	2	G81114	outer membrane pro
43	87	9.5	21.2	2	A82212	opacity protein-re
44	86.5	9.5	350	2	A82299	outer membrane pro
45	86	9.4	394	2	S34263	outer membrane por

## ALIGNMENTS

```

RESULT      1
F64124
opacity protein homolog H11457 - Haemophilus influenzae (strain Rd kw20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: F64124
R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirnness, E.F.; Kerlavage,
  Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Wetnam
  , D.M.; Brannon, R.C.; Fine, L.D.; Fritchman, J.L.; Furumann, J.L.; Geoghagen, N.S.M.
  Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
  A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
  A:Reference number: A64000; MUID:95350630
A:Accession: F64124
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-178 <TIGR>
A:Cross-references: GB:U02823; GB:L42023; NID:g1574281; PIDN:AC23104.1; PID:g1574294

Query Match      13.5%; Score 123.5; DB 2; Length 178;
Best Local Similarity 27.5%; Pred. No. 0.0017;
Matches 52; Conservative 30; Mismatches 82; Indels 25; Gaps 8;

QY 4 IKTLLAVSASSLLANSANAISYGSNADAPQYVAGKIG--QVDAKQINGKMT--AYGIY 58
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MKKLLITVMLFTLAL-----SAQAQWYVQGLDASKRIDITHVNSSNSPFTQHS 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 59 AGYNNDONFGVYAEVEVSGDAKEFN-AGVSPYKGDVKSFGCAGTYYNTINPFAKKG 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 51 VGYAPDKNERLAVDPTNYGKYANADYDVVSLKRSGLTGDFDLADPKRPYV--GVR 108
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 118 IAKTIVDVTSRAATPYYSKSKDKTSLAGV---GYGFRPLANVGVAS--VNYSESDANA 171
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 109 VSTNCADYTA-NARYRIEAFETFRIGIGALAGVQYKLTQVNLNTIEYRLASVSD 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 172 ISLGAHLAF 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 168 VGKAGLRF 176
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT      2
A35123
all protein precursor - Yersinia enterocolitica
C:Species: Yersinia enterocolitica
C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 02-Sep-2000
C:Accession: A35123
R:Miller, V.L.; Bliska, J.B.; Falkow, S.
  J. Bacteriol. 172, 1062-1069, 1990
A:Title: Nucleotide sequence of the Yersinia enterocolitica All gene and characteriza
  A:Reference number: A35123; MUID:90130261
A:Accession: A35123

```

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-178 <MIL>  
A:Cross-references: GB:M29945; NID:g155437; PIDN:AAA8694.1; PID:g155438  
C:Superfamily: phage lambda membrane protein lom

Query Match 13.5%; Score 123; DB 2; Length 178;  
Best Local Similarity 29.6%; Pred. No. 0.0018;  
Matches 55; Conservative 29; Mismatches 82; Indels 20; Gaps 9;

QY 5 KTLAVSASSLLAMSANAISYGSNADAPYVAKIGOV--DAKQINCKNTAYGIYAGYN 62  
DB 3 KTLAL---ASSLIACLSISVNYTAASESSISIGVASHKKEGTYLDNDPKFNLKYRTE 59  
QY 63 FDQNGVEAEFVGS-DAKEFNAGVSPV-KGDVKSFGAYGYRNF-INTPFYAKKLGIA 119  
DB 60 LDDMGVIGSFAYTHQGVDFYGSNKFHGGVDYYSV--TMCPSRINRYVSLYGLGMA 117  
QY 120 KTKVDVTSRNATY--SNKSDKTSIAGVGVGFKPLANVGVEASTNYLSEDAHAIS---L 174  
DB 118 HCKV-----KASVFDESSISASKTSMAYGAGVQFNPLPFVIDASYEYSKLDSIKYGTWL 172  
QY 175 GAHLAF 180  
DB 173 GAGYRF 178

RESULT 3  
H81277  
flagellin CJ1339c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: H81277  
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chilling  
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrer  
Nature 403, 665-668, 2000  
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hly  
A:Reference number: A81250; MUID:20150912  
A:Accession: H81277  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-572 <PAR>  
A:Cross-references: GB:A1139078; GB:A1111168; NID:g6968723; PIDN:CAB73766.1; PID:g696877  
A:Experimental source: serotype O2, strain NCTC 11168  
C:Genetics:  
A:Gene: flaA; CJ1339c  
C:Superfamily: flagellin

Query Match 12.8%; Score 117; DB 2; Length 572;  
Best Local Similarity 24.5%; Pred. No. 0.022;  
Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AAIYSNSADAPYVAKIGOVDAKO-----INCKNTAYGIYA----- 59  
DB 242 AAVRAGAVSDFAINGVIGKVDYDGDANGALVAIINSVKDTTVEASIDANGQLLITS 301  
QY 60 ---GYNPDQNGVEAEFVGSDAKEFNAGVSPVKGDV-----SEFAGTYRY 103  
DB 302 REGRIKIKIDNIGCGA-FINADMKENYGRSLVKKNDGKDLILSGSNLSAGCA----- 354  
QY 104 NEINPFYAKKGLGIATK--VDVTSRNATYYSNKSDDKTSIAG-----GAVGVGFKP 152  
DB 355 ---TQFISQAVSLRESKQIDANIADAMGRHSANKGVGLGYSVSAVMSASGSFSS 410  
QY 153 LANVGVEASTNYLSPDANAISLGA 176  
DB 411 GSGYSGVSGKNYSTGFANAAISA 434

RESULT 4  
H82207

hypothetical protein VC1384 [imported] - Vibrio cholerae (group O1 strain N16961)  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 15-Sep-2000  
C:Accession: H82207  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.  
chardson, D.; Ermolaeva, M.D.; Vamthyan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: H82207

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-160 <HET>  
A:Cross-references: GB:AE004217; GB:AE003852; NID:g9655866; PIDN:AAF94542.1; GSPDB:GN  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:  
A:Gene: VC1384  
A:Map position: 1

Query Match 12.6%; Score 114.5; DB 2; Length 160;  
Best Local Similarity 29.8%; Pred. No. 0.0084;  
Matches 56; Conservative 19; Mismatches 70; Indels 43; Gaps 9;

QY 5 KTLAVSASSLLAMSANAISYGSNADAPYVAKIGOVDAKQINCKNTAYGIYAGYND 64  
DB 3 KTLAL---ALLGASSTA-----MADSMYIGAGVAGSGSDYEGKHG--TAVSVHAGTIL 51  
QY 65 QNFGEAEFVGSDAKEFNA-----GVSYPKGDVKSNGAAGTYRYNINPFYAKKL----- 116  
DB 52 PTGLEAGVYVNHDFEINATQELASSLYFAKPSMDGPLH-----VYAKGGLHSD 104  
QY 117 -GIATKVDVTSRNATYYSNKSDDKTSIAGVGVGFKPLANVGVEASY--NYL--SEDAHA 172  
DB 105 KQINCKRID-----DGLDVMYGIAGAEFIILGPRSVGASVYNTMTSDVGTIL 151  
QY 173 SLGAHLAF 180  
DB 152 SFNATPHF 159

RESULT 5  
D64050  
major outer membrane protein P2 - Haemophilus influenzae (strain Rd Kw20)

C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 21-Jul-2000  
C:Accession: D64050  
R:Feilschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kertlavage  
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vence  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: D64050  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-359 <TIGR>  
A:Cross-references: GB:U22699; GB:U42023; NID:g9212180; PIDN:AAC21810.1; PID:g1573092  
C:Keywords: membrane protein

Query Match 12.3%; Score 112.5; DB 2; Length 359;  
Best Local Similarity 28.1%; Pred. No. 0.031;  
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY 5 KTLAVSASSLLAMSANAISYGSNADAPYVAKIGOV-----DAKQINCKNTAYGIYAGYND 64  
DB 3 KTLALIVGAFAPASANAAYVYNNCG-TNVELGRLSLIAQGSNTVNOQOQALANQ 61  
QY 52 NTAYGIYAGYNDQNF-----GYEAEFY-----GSDAKEFNAGVSPVKGDVSGFAGTY 101

```

Db      62  GSREFIKATNHFEGDGFVAQGYLETFRFVTKASENSD--NF-----GDITS-----Ky 106
QY      102  RYNFNINPFYKAGKLGIAKTVKD--VTSRNATTVS--NKSDDTSIAGC-VGVGFPLANVG 157
      1 : 111 111 : 111 : 111 : 111 :
Db      107  AYVVLGNKAFGEVFLGKRAKTIADGITSAEDEKGYGLNNSDIPTPSGNTVGYTFGIDGLV 166
QY      158  VEASY 162
      158 : 1-1
Db      167  LGANV 171

RESULT      6
A30542
major outer membrane protein P2 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 08-Oct-1999
C:Accession: A30542; A32750; A60116; A60121; S09622
R:Monson Jr., R.; Tojan Jr., R.W.
Infct. Immun. 57, 88-94, 1989
A:Title: Molecular cloning, expression, and primary sequence of outer membrane protein P2
A:Reference number: A30542; MUID:89079316
A:Accession: A30542
A:Molecule type: DNA
A:Residues: 1-361 <MUN>
A:Cross-references: GB:003359; NID:g148958; PIDN:AAA24993.1; PID:g148959
A:Experimental source: type b
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed
R:Hansen, E.J.; Hasegama, C.; Clausell, A.; Capra, J.D.; Orth, K.; Moomaw, C.R.; Slaughter
Infct. Immun. 57, 1100-1107, 1989
A:Title: Primary structure of the porin protein of Haemophilus influenzae type b determined
A:Reference number: A32750; MUID:89173305
A:Accession: A32750
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-361 <HAN>
A:Experimental source: type b, strain DL42
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed
R:Hansen, E.J.; Pelzel, S.E.; Orth, K.; Moomaw, C.R.; Radolf, J.D.; Slaughter, C.A.
Infct. Immun. 57, 3270-3275, 1989
A:Title: Structural and antigenic conservation of the P2 porin protein among strains of
A:Reference number: A60116; MUID:90035390
A:Accession: A60116
A:Molecule type: protein
A:Residues: 21-42 <HA2>
A:Experimental source: type b, strains DL42, CH100, DL26, DV102, H234, Madigan, OA104, and
R:Hansen, E.J.; Gonzales, F.R.; Chamberlain, N.R.; Norgard, M.V.; Miller, E.E.; Cope, L.
Infct. Immun. 56, 2709-2716, 1988
A:Title: Cloning of the gene encoding the major outer membrane protein of Haemophilus influenzae
A:Reference number: A60121; MUID:86330197
A:Accession: A60121
A:Molecule type: protein
A:Residues: 21-40 <HA3>
A:Experimental source: type b, strain DL42
R:Monson Jr., R.; Bailey, C.; Grass, S.
Mol. Microbiol. 3, 1797-1803, 1989
A:Title: Diversity of the outer membrane protein P2 gene from major clones of Haemophilus influenzae
A:Reference number: S09622; MUID:90158127
A:Accession: S09622
A:Status: preliminary: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-361 <MU2>
A:Keywords: membrane protein
F:1-20/Domains: signal sequence #status predicted <SIG>
F:21-361/Product: major outer membrane protein P2 #status experimental <MAT>

Query Match      12.3%, Score 112.5; DB 2; Length 361;
Best Local Similarity 28.1%, Pred. No. 0.031;
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;

QY      5  KTLAVASASLLAMSAANAISGNSADQAPYVGATIGV-----DAQING-----K 51
      11 : 111 111 : 111 : 111 : 111 :
Db      3  KTLAALIVGAFAPASAAANAAYVYNNEG-TNVELGAGLSITADQSNSTVDNKOCHOALRNO 61

```

OY 52 NTAGIVGVYGNEDQNF-----GVEAEFV-----GSDAEFNAGVSPYKGDVKSFGANGTY 101  
 Db 62 GSRFIKATKTHNFGCGFYAGQYLETRFVTKASENSD--NF-----GDITTS-----KY 106  
 OY 102 RYNFINTFFYAKGKLGIAKTRVD--VTSRNATTYS--NKSDDKTSIAGC--VGVGFKPLANVG 157  
 Db 107 AYVTLGNKAFAEGEVKLGRAKRTIADGITSADREXVGLNNSDYIPIPSGNTVGYTFKIGDGLV 166  
 OY 158 YEAST 162  
 Db 167 LGANY 171  
 RESULT 7  
 154668  
 heat resistant agglutinin 1 - Escherichia coli  
 C:Species: Escherichia coli  
 C:Date: 07-Jun-1996 #sequence\_revision 07-Jun-1996 #text\_change 08-Oct-1999  
 C:Accession: 154668; 169133; 169134  
 R:Lutwyche, P.; Ruppel, R.; Cavanagh, J.; Warren, R.A.; Brooks, D.E.  
 Infect. Immun. 62, 5020-5026, 1994  
 A:Title: Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin from *Escherichia coli* O157:H7  
 A:Reference number: 154668; MUID:95012721  
 A:Accession: 154668  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-264 <RES>  
 A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AAIC13751.1; PID:g463911  
 A:Accession: 169133  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 15-264 <RE2>  
 A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AAIC13752.1; PID:g463912  
 A:Accession: 169134  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 18-264 <RE3>  
 A:Cross-references: EMBL:U07174; NID:g463910; PIDN:AAIC13753.1; PID:g463913  
 Query Match 12.0%; Score 109.5; DB 2; Length 264;  
 Best Local Similarity 21.0%; Pred. No. 0.039;  
 Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;  
 OY 1 MKTLLTLLAVNSSSLILANSANNAIS-----YGSADAQP 34  
 Db 15 MLENKRVIALVSALAMAGNFSTQALADESKTGFYTGKAGASVMSIADQRFSLSGNEETSK 74  
 OY 35 YVGAKIGQVDAAKOIIGKMTAY--GIYAGVYNPQNGV----EAEFV--GSDAKFENAGVS 86  
 Db 75 YKGG-----DGHDTVFSGGIAAGYDTPQPSIVRKELEFVYARAKADSKYINVKD 124  
 OY 87 PVKG----DVKS-----FGAYGYRYNFINTFFYAKKLGIAK-----TKVDV-- 125  
 Db 125 SMSGSGWMDKLDNEYSVNTLMNLNAVYDERNSAPFPPWSAG--IGYAKKEIHQKTGISTWD 183  
 OY 126 -----TSRNATTYSNKSDDKTSIAGCGVGFKPLANVGAEASNYLSEDAANNAIS 173  
 Db 184 YGYSSGREGSLSRSGSADNFAMWSLGAVGRYDVTPIADLSTRYLDADGSSVS 236  
 RESULT 8  
 A39185  
 virulence protein pagC precursor - Salmonella typhimurium  
 C:Species: Salmonella typhimurium  
 C:Date: 08-Nov-1991 #sequence\_revision 08-Nov-1991 #text\_change 02-Sep-2000  
 C:Accession: A39185  
 R:Pulkkinen, W.S.; Miller, S.I.  
 J. Bacteriol. 173, 86-93, 1991  
 A:Title: A Salmonella typhimurium virulence protein is similar to a *Yersinia enterocolica* enterocolicin  
 A:Reference number: A39185; MUID:911100323  
 A:Accession: A39185

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-188 <PUI>  
A:Cross-references: GB:M5546; NID:g154232; PIDN:AAA27179.1; PID:g154233  
C:Superfamily: phage lambda membrane protein Iom

Query Match 11.8%; Score 108; DB 2; Length 188;  
Best Local Similarity 24.2%; Pred. No. 0.036;  
Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;

OY 1 MKTLTLLAVSSSLASANAISYGSADQPVYAKIGGVDAKOTNGKRTATGIYAG 60  
DB 1 MKNITLSTLVTTTSLVYVNAQADTNAFSVGYARKAQSKV--QDERKNIKGVNWKY----- 53  
OY 61 YNFDNFGVEAFVGSDAKEFNAGVSPYKGDVKSFGA-----YGT----- 100  
DB 54 -----RYEDDSPVSFISLSLYLXGDRQAGSVPEGSIYHDKFEVYKGSIMWGP 102  
OY 101 -YRYFINTPFAKAKGLIAK-TKVDVTSRNAITYSNK--SDKTSLAGVGVEKPLANV 156  
DB 103 AVRLS-DNFSLYALAGVGVKATPEKHSYODGDSFNKISSRKTGFAMCAGVQNMPLNI 161  
OY 157 GVEASY---NYLSEDAANISLGAHLAF 180  
DB 162 VVDVGIEGNSISSTKINGFNVGVTFR 188

RESULT 9  
MMBPP2  
outer membrane porin lc precursor - phage PA2  
C:Species: phage PA2  
A:Note: host Escherichia coli  
C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 02-Jul-1998  
C:Accession: D25647  
R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.  
J. Biol. Chem. 261, 12723-12732, 1986  
A:Title: Structure of the lc and mmpc outer membrane porin protein genes of lambdaoid bac  
A:Reference number: A25647; MUID:86304457  
A:Accession: D25647  
A:Molecule type: DNA  
A:Residues: 1-365 <BLA>  
C:Genetics:  
A:Gene: lc  
C:Superfamily: outer membrane protein phoe  
C:Keywords: membrane protein; porin; trimer  
F:1-23/Domain: signal sequence #status predicted <SIG>  
F:24-365/Product: outer membrane porin lc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;  
Best Local Similarity 26.6%; Pred. No. 0.076;  
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;  
OY 1 MKTLTLL-AVSASSLLMSANAISY-----GNSADAQPYVAK 39  
DB 1 MKKLVAISAVAASVLMASQAALTYNKDSNKLDLYGVNAKHVSSNDADDGDTTYAR 60  
OY 40 IGOVDAKINGKNTFYAGYAGYFNPDQNGVEAEFVGSDAKEFNAGVSPYKGDVKSFGAYG 99  
DB 61 LGFKETQINDQLTGFQGW-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111  
OY 100 TYRYNFINTPFAKAKL-----GIATKTVDV-----TSRN----- 129  
DB 112 SIDVGRNVAVDIGAWTDVLPFGGDTWTQTDVEMTGRITGFATYRRNDDPGLVDGLNF 171  
OY 130 ATTYSNKSDKT---SLAGVGVEKPLANVGEVASYNTLSGDANAISLGAHLA 179  
DB 172 AAQYOGKNDKSDPDMYTEGNGDGF-----GFSATYEV-----EGFGIGATYA 213

RESULT 10  
MMBCKC

outer membrane porin mmpc precursor - Escherichia coli cryptic lambdaoid prophage DLP1

C:Species: Escherichia coli  
C>Date: 31-Dec-1990 #sequence\_revision 31-Dec-1990 #text\_change 16-Jul-1999  
C:Accession: A25647; S66594; G64787

R:Blasband, A.J.; Marcotte Jr., W.R.; Schnaitman, C.A.

J. Biol. Chem. 261, 12723-12732, 1986

A:Title: Structure of the lc and mmpc outer membrane porin protein genes of lambdaoid

A:Reference number: A25647; MUID:86304457

A:Contents: mutant strain CS384

A:Accession: A25647

A:Molecule type: DNA

A:Residues: 1-365 <BLA>

R:Mandl, A.A.; Sharples, G.J.; Mandal, T.N.; Lloyd, R.G.

J. Mol. Biol. 257, 561-573, 1996

A:Title: Holliday junction resolvases encoded by homologous rusa genes in Escherichia

A:Reference number: S66579; MUID:96196428

A:Accession: S66594

A:Molecule type: DNA

A:Residues: 347-365 <MAH>

A:Cross-references: EMBL:X92587; NID:g1051136; PIDN:CAA63325.1; PID:g1051145

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G64787

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 'NNIYRAVTSFENNSSKKGLT',1-325, 'N',327-347, 'EGAKSKSLT' <BLAT>

A:Cross-references: GB:AE000160; GB:U00096; NID:g1786751; PIDN:AACT73654.1; PID:g17867

A:Experimental source: Strain K-12, Substrain W61655

C:Comment: In wild-type strains of E. coli K-12, the mmpc open reading frame is inter

ain CS348, the IS5 element has been deleted and mmpc is expressed.

C:Genetics:

A:Gene: mmpc

A:Map position: 12 min

A:Genome: cryptic lambdaoid prophage DLP12

C:Superfamily: outer membrane protein phoe

C:Keywords: membrane protein; porin; trimer

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-365/Product: outer membrane porin mmpc #status predicted <MAT>

Query Match 11.8%; Score 108; DB 1; Length 365;  
Best Local Similarity 26.6%; Pred. No. 0.076;  
Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;  
OY 1 MKTLTLL-AVSASSLLMSANAISY-----GNSADAQPYVAK 39  
DB 1 MKKLVAISAVAASVLMASQAALTYNKDSNKLDLYGVNAKHVSSNDADDGDTTYAR 60  
OY 40 IGOVDAKINGKNTFYAGYAGYFNPDQNGVEAEFVGSDAKEFNAGVSPYKGDVKSFGAYG 99  
DB 61 LGFKETQINDQLTGFQGW-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111  
OY 100 TYRYNFINTPFAKAKL-----GIATKTVDV-----TSRN----- 129  
DB 112 SIDVGRNVAVDIGAWTDVLPFGGDTWTQTDVEMTGRATGVATYRRNDDPGLVDGLNF 171  
OY 130 ATTYSNKSDKT---SLAGVGVEKPLANVGEVASYNTLSGDANAISLGAHLA 179  
DB 172 AAQYOGKNDKSDPDMYTEGNGDGF-----GFSATYEV-----EGFGIGATYA 213

RESULT 11  
S68072  
major outer membrane protein P2, antigenic variant t1 - Haemophilus influenzae  
C:Species: Haemophilus influenzae  
C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 08-Oct-1999  
C:Accession: S68072; S45437  
R:Dunm, B.  
submitted to the EMBL Data Library, June 1993  
A:Reference number: S68062







GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 6, 2001, 14:40:04 ; Search time 14.8 seconds  
(without alignments)  
392.766 Million cell updates/sec

Title: US-09-164-714-7  
Perfect score: 912  
Sequence: 1 MKTILKTLAVSASSLLAMSA.....SYNYLSEDAVAISGAHLAP 180

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	123.5	13.5	178	1	YES7_HAEIN
2	123	13.5	178	1	ATL1_TREN
3	117	12.8	571	1	FLN1_CAMJE
4	115.5	12.7	182	1	ATL1_YERPS
5	113	12.4	359	1	OM53_HAEIN
6	112.5	12.3	359	1	OM21_HAEIN
7	112.5	12.3	361	1	OM22_HAEIN
8	109	12.0	365	1	POR1_BPPA2
9	108	11.8	188	1	PACQ_SALTY
10	108	11.8	365	1	NMPC_ECOLI
11	105	11.5	240	1	OM31_BRUME
12	105	11.5	363	1	OM2B_HAEIN
13	103	11.3	371	1	OM25_HAEIN
14	102.5	11.2	213	1	OM25_BRUAB
15	102.5	11.2	1310	1	VAC3_HELPY
16	101	11.1	386	1	OM24_HAEIN
17	101	11.1	571	1	FLN1_CAMJE
18	100.5	11.0	201	1	OM25_BRUOY
19	100	11.0	353	1	OM52_HAEIN
20	100	11.0	361	1	OM29_HAEIN
21	100	11.0	1287	1	VAC2_HELPY
22	100	11.0	1291	1	VAC4_HELPY
23	99.5	10.9	172	1	OMPX_ENTCL
24	99.5	10.9	367	1	OM2A_HAEIN
25	99	10.9	360	1	OM26_HAEIN
26	98.5	10.8	213	1	OM25_BRUNE
27	98.5	10.8	369	1	OM27_HAEIN
28	96	10.5	385	1	OM23_HAEIN
29	95.5	10.5	1290	1	VAC0_HELPY
30	95	10.4	211	1	ROBP_RHIVY
31	95	10.4	574	1	FLA3_CAMJE
32	94.5	10.4	213	1	OM25_BRUCA
33	94.5	10.4	213	1	OM25_BRUME

34	94.5	10.4	213	1	OM25_BRUSU	Q45689 bruceella su
35	90	9.9	574	1	FLB3_CAMJE	Q46114 camylobact
36	90	9.9	1296	1	VAC1_HELPY	Q48247 helicobacte
37	89.5	9.8	1585	1	YOBO_BACSU	P45931 bacillus su
38	89	9.8	1771	1	OMPX_ECOLI	P36546 escherichia
39	89	9.8	359	1	OMPA_SERMA	P04845 serratia ma
40	88	9.6	346	1	OMPA_ECOLI	P02934 escherichia
41	86	9.4	394	1	OM51_SALTI	Q56110 salmonella
42	86	9.4	623	1	VPS_HAEIN	P44242 haemophilus
43	85.5	9.4	341	1	OMPU_VIRCH	P97085 vibrio chol
44	85.5	9.4	356	1	VATX_DICDI	P54641 dictyostell
45	84.5	9.3	351	1	OMPA_SHIDY	P02935 shigella dy

## ALIGNMENTS

```

RESULT 1
ID YES7_HAEIN STANDARD; PRT; 178 AA.
AC 057201: 005063:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN H11457 PRECURSOR.
GN H11457.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20:
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Uetelman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Wetterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC - SIMILARITY: WEAK. TO N.CONORRHOAE OPACITY PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U32823: AAC23104.1: -.
CC DR TIGR: H11457: -.
CC KW Hypothetical protein; Signal.
CC FT SIGNAL 1
CC CHAIN 20
CC SEQUENCE 178 AA; 19470 MW; 150AC5C44FF49AFF CRC64;

```

Query Match 13.5%; Score 123.5; DB 1; Length 178;

Best local Similarity 27.5%; Pred. No. 0.00083;

Matches 52; Conservative 30; Mismatches 82; Indels 25; Gaps 8;

QY 4 LKTLAVSASSLLAMSAISGNADQPYGAKIG--QVDAKQINKNT---AVGY 58  
: | | | : | | | | : | | | : | | | : | | | :  
DB 1 MKKLLIVTMFLTAL-----SAQAWYVQGLGASKIDITHVNSNSGSPFQRLS 50  
: | | | : | | | : | | | : | | | : | | | :  
QY 59 AGYNFDQNGVEAEFVGSDAKFEFN-AGVSPVKGQVKSFGAYGTYRNYINTPFYAKGKLG 117  
: | | | : | | : | | | : | | | : | | | : | | | :

```
Db 51 VGYAFDKNRFLAVDTYNGKVTANADVVDYSLKSGSLGTGFDPDLADPKPY--GVR 108
OY 118 IAKTVVDYTSRNATYYSKSKDTSIAGV---GVGEKPLANVGEAS--YNYISEDANA 171
Db 109 VSTNCADYTA-NARYRIEAFETETRICIGALAGVQYKLTQVVAINTNIEYRLASNVSD 167
OY 172 ISLGAHLAF 180
Db 168 VGKAGLRF 176

RESULT 2
AII_YEREN
ID AII_YEREN STANDARD: PRT: 178 AA.
AC P16454;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ATTACHMENT INVASION LOCUS PROTEIN PRECURSOR.
GN AII.
OS Yersinia enterocolitica.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Yersinia.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-33.
RX MEDLINE-90130261; PubMed-168838;
RA Miller V.L., Bliska J.B., Falkow S.;
RT "Nucleotide sequence of the Yersinia enterocolitica aII gene and
RT characterization of the AII protein product.";
RL J. Bacteriol. 172:1062-1069(1990).
CC -I- FUNCTION: THIS MEMBRANE-ASSOCIATED PROTEIN PROMOTES INVASION
CC OF PATHOGENIC BACTERIA INTO EUKARYOTIC CELLS BY AN UNKNOWN
CC MECHANISM.
CC -I- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -I- SIMILARITY: BELONGS TO THE AII/OMPX/PACG/LDM FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, M29945; AAA8694.1; -.
CC PIR, A5123; A5123.
CC INTERPRO: IPR000758; -.
CC PRINTS: PR00316; ENTEROVIROMP.
CC PROSITE: PS00694; ENT_VIR_OMP_1; 1.
CC PROSITE: PS00695; ENT_VIR_OMP_2; 1.
CC Outer membrane; Transmembrane; Signal; Virulence.
CC SIGNAL 1 23
CC CHAIN 24 178 ATTACHMENT INVASION LOCUS PROTEIN.
CC FT 178 AA; 19548 MW; D46FC92E2524F162 CRC64;
CC SEQUENCE

Query Match 13.5%, Score 123; DB 1; Length 178;
Best Local Similarity 29.6%; Pred. No. 0.00091;
Matches 55; Conservative 29; Mismatches 82; Indels 20; Gaps 9;
```

```
Db 173 GAGYRF 178

RESULT 3
FLAI_CAMJE
ID FLAI_CAMJE STANDARD: PRT: 571 AA.
AC P56963; Q9PMW0;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FLAGELLIN A.
GN FLA A OR C11339C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
CC Campylobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCTC 11168;
RX MEDLINE-20150912; PubMed-10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Felwell T., Holtroyd S.,
RA Jagers K., Kariyasev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Randalam M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrett B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -I- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -I- SUBUNIT: HETEROPOLYMER OF FLA A AND FLA B.
CC -I- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AL139078; CAB3766.1; -.
CC KW Flagella.
CC FT INIT-MET 0 0 BY SIMILARITY.
CC FT SEQUENCE 571 AA; 58907 MW; 9C63B2A10C1AE863 CRC64;
CC SEQUENCE

Query Match 12.8%, Score 117; DB 1; Length 571;
Best Local Similarity 24.5%; Pred. No. 0.01;
Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;
```



RA MEDLINE-95350630; PubMed-7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shiley R., Liu L.-I., Glodet A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uitterlbeck T.R., Hanna M.C., Nguyen D.T., Saudok D.M., Brandon R.C.,  
RA Fine L.D., Fitchman J.L., Fuhmann J.L., Georgiades N.S.M.,  
RA Genchik C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
RA Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT Influenzae Rd.":  
RL Science 269:496-512(1995).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U32699; AAC21810.1; -  
DR TIGR: H10139; -  
KW Outer membrane; Transmembrane; Porin; Signal.  
KW SIGNAL 1 20 BY SIMILARITY.  
FT CHAIN 21 359 OUTER MEMBRANE PROTEIN P2.  
SQ SEQUENCE 359 AA; 39375 MW; 7E9E9FEC2BD50314 CRC64;  
  
Query Match 12.3%; Score 112.5; DB 1; Length 359;  
Best Local Similarity 28.1%; Pred. No. 0.015;  
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;  
  
QY 5 KTLAVSASSLLAMANAISYSGNSADAPYVGAKIGOV-----DAKQING-----K 51  
DB 3 KTLAALIYGAFAASAANAAYVYNNEG-TNVELGGRSLTIAEOSNSTVDNOKOQHGALRNQ 61  
QY 52 NTAAGIYAGYNPDQNF---GVEAEFV-----GSDAKEFNAGVSPVKGDVKSFGAGCTY 101  
DB 62 GSRFHAKATHNFGDGFYAGYLETREFTVKASNGSD--NF-----GDITS-----KY 106  
QY 102 RYNEINTPEYAKGLGIATKYD-VTSRNATYTS--NKSDDKSLAGG-VGVGFKPLANVG 157  
DB 107 AVYTLGNKNAFGVEKLGRAKTIADGITSADKEKEYGVLYNSDYIPTSGNTVGYTFKIDGLV 166  
QY 158 YEAST 162  
DB 167 LGANY 171  
  
RESULT 7  
OM22\_HAEIN STANDARD; PRT; 361 AA.  
ID OM22\_HAEIN STANDARD; PRT; 361 AA.  
AC P20149;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 01-NOV-1995 (Rel. 32, Last annotation update)  
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).  
GN OMP2.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SEROTYPE B;  
RX MEDLINE-89079316; PubMed-2535836;  
RT Munson R.S. Jr., Tolian R.W. Jr.;  
RT "Molecular cloning, expression, and primary sequence of outer  
RT membrane protein P2 of Haemophilus influenzae type b.";  
RL Infect. Immun. 57:88-94(1989).

RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SEROTYPE B;  
RX MEDLINE-90158127; PubMed-2576096;  
RA Munson R.S. Jr., Bailey C., Grass S.;  
RT "Diversity of the outer membrane protein P2 gene from major clones of  
RT Haemophilus influenzae type b";  
RL Mol. Microbiol. 3:1797-1803(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SEROTYPE B;  
RX MEDLINE-89173305; PubMed-2538396;  
RA Hansen E.J., Hasemann C., Clausell A., Capra J.D., Orth K.,  
RA Moomaw C.R., Slaughter C.A., Latimer J.D., Miller E.E.;  
RT "Primary structure of the porin protein of Haemophilus influenzae  
RT type b determined by nucleotide sequence analysis";  
RL Infect. Immun. 57:1100-1107(1989).  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
CC -1- SIMILARITY: TO PHOTOBACTERIUM STRAIN SS9 OMPH.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: J03359; AAA24993.1; -  
DR EMBL: A09003; CA00819.1; -  
DR PIR: A30542; A30542.  
DR PIR: S09622; S09622.  
KW Outer membrane; Transmembrane; Porin; Signal.  
KW SIGNAL 1 20  
FT CHAIN 21 361 OUTER MEMBRANE PROTEIN P2.  
SQ SEQUENCE 361 AA; 39701 MW; AF104C9CC8942D69 CRC64;  
  
Query Match 12.3%; Score 112.5; DB 1; Length 361;  
Best Local Similarity 28.1%; Pred. No. 0.015;  
Matches 52; Conservative 25; Mismatches 65; Indels 43; Gaps 11;  
  
QY 5 KTLAVSASSLLAMANAISYSGNSADAPYVGAKIGOV-----DAKQING-----K 51  
DB 3 KTLAALIYGAFAASAANAAYVYNNEG-TNVELGGRSLTIAEOSNSTVDNOKOQHGALRNQ 61  
QY 52 NTAAGIYAGYNPDQNF---GVEAEFV-----GSDAKEFNAGVSPVKGDVKSFGAGCTY 101  
DB 62 GSRFHAKATHNFGDGFYAGYLETREFTVKASNGSD--NF-----GDITS-----KY 106  
QY 102 RYNEINTPEYAKGLGIATKYD-VTSRNATYTS--NKSDDKSLAGG-VGVGFKPLANVG 157  
DB 107 AVYTLGNKNAFGVEKLGRAKTIADGITSADKEKEYGVLYNSDYIPTSGNTVGYTFKIDGLV 166  
QY 158 YEAST 162  
DB 167 LGANY 171  
  
RESULT 8  
PORI\_BPPA2 STANDARD; PRT; 365 AA.  
ID PORI\_BPPA2 STANDARD; PRT; 365 AA.  
AC P07238;  
DT 01-APR-1988 (Rel. 07, Created)  
DT 01-APR-1988 (Rel. 07, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE OUTER MEMBRANE PORIN PROTEIN LC PRECURSOR.  
GN LC.  
OS Bacteriophage PA-2.  
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-86304457; PubMed-3017988;

RA Blasband A.J., Marcotte W.R. Jr., Schnaitman C.A.;  
 RT "Structure of the 1c and nmcp outer membrane porin protein genes of  
 RT lambdaoid bacteriophage."  
 RT J. Biol. Chem. 261:12723-12732(1986).  
 CC -1- FUNCTION: PORINS ARE MAJOR PROTEINS FOUND IN THE OUTER MEMBRANES  
 CC OF GRAM-NEGATIVE BACTERIA WHERE THEY FORM CHANNELS FOR THE  
 CC NONSPECIFIC PERMEATION OF SMALL SOLUTES (MOLECULES WITH MW LOWER  
 CC THAN 4000-6000 DALTONS).  
 CC -1- SUBUNIT: HOMOTRIMER (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE  
 CC (BY SIMILARITY).  
 CC -1- MISCELLANEOUS: A PORIN GENE CAN ALSO BE FOUND IN THE GENOMES OF  
 CC CERTAIN LAMBDOID BACTERIOPHAGE, AND ITS PROTEIN IS EXPRESSED IN  
 CC THE LYSOGENIC STATE. IN E. COLI THE EXPRESSION OF OMPC AND OMPE  
 CC PROTEINS IS THEN REDUCED SUBSTANTIALLY.  
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: J02580; AAA32301.1; -.  
 DR PIR: D25647; MMBP2.  
 DR HSSP: P02931; IGFN.  
 DR INTERPRO: IPR001702; -.  
 DR INTERPRO: IPR001897; -.  
 DR PFAM: PF00267; Gram-ve\_porins; 1.  
 DR PRINTS: PR00182; ECOLI\_PORIN.  
 DR PRINTS: PR00183; ECOLI\_PORIN.  
 DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
 KM Outer membrane; Transmembrane; Porin; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN LC.  
 FT CONFLICT 99 99 H -> R (IN AA SEQUENCE).  
 FT SEQUENCE 365 AA; 40290 MW; 0FBC0531FB9C0205 CRC64;  
 SQ  
 Query Match 12.0%; Score 109; DB 1; Length 365;  
 Best Local Similarity 26.6%; Pred. No. 0.029;  
 Matches 62; Conservative 20; Mismatches 77; Indels 74; Gaps 11;  
 QY 1 MKTLKTL-AYASASSLLASANAISY-----GNSADAQPIYVGA 39  
 DB 1 MKRLVAISAVAASVLAASQAALTYNKRDSNKLDTYGVNAKHYFSSDPADGDTTYAR 60  
 QY 40 IGVDAKQINGKNTAVGIAGYNFDQNFGEAEFVGSDAKERNAGVSPKGDVKSFGAVG 99  
 DB 61 LQFKGTQINDQITGEGWE-YEFKGN---RAESQSSSKDKTHLAFAGLK-----FGDYG 111  
 QY 100 TYRYNFINTPFYAKGKL-----GIATKYDV-----TSRN----- 129  
 DB 112 SIDYGRNNGVAVDIDGAMTDVLDPEFGDVTWQIDVFWTGRRTGATFAYRNDFGLVDGLNF 171  
 QY 130 ATTSYKSKSKT---SLAGVGVGFKPLANVGVEASYNISYSEDANAISLGAHLA 179  
 DB 172 AAQYQKNDKSDPDNYTEGNGDGF-----GFSATVEY-----EGFGIGATYA 213  
 RESULT 9  
 PAGC\_SALTY STANDARD; PRT; 188 AA.  
 AC P23988;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE VIRULENCE MEMBRANE PROTEIN PAGC PRECURSOR.  
 GN PAGC.  
 OS Salmonella typhimurium.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.  
 RP [1]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN-ATCC 14028;  
 RX MEDLINE-91100323; PubMed-1846140;  
 RA Pulkkinen W.S., Miller S.L.;  
 RT "A Salmonella typhimurium virulence protein is similar to a Yersinia  
 RT enterocolitica invasion protein and a bacteriophage lambda outer  
 RT membrane protein."  
 RT J. Bacteriol. 173:86-93(1991).  
 CC -1- FUNCTION: ESSENTIAL FOR FULL VIRULENCE AND SURVIVAL WITHIN  
 CC MACROPHAGES.  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE AIL/OMPX/PAGC/LOW FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M55546; AAA27179.1; -.  
 DR PIR: A39185; A39185.  
 DR STGENE: SG10676; PAGC.  
 DR INTERPRO: IPR000758; -.  
 DR PRINTS: PR00316; ENTEROVIROMP.  
 DR PROSITE: PS00694; ENT\_VIR\_OMP\_1; 1.  
 DR PROSITE: PS00695; ENT\_VIR\_OMP\_2; 1.  
 KM Outer membrane; Transmembrane; Signal; Virulence.  
 FT SIGNAL 1 23  
 FT CHAIN 24 188 VIRULENCE MEMBRANE PROTEIN PAGC.  
 FT SEQUENCE 188 AA; 20574 MW; B23826F42B62DDFE CRC64;  
 SQ  
 Query Match 11.8%; Score 108; DB 1; Length 188;  
 Best Local Similarity 24.2%; Pred. No. 0.017;  
 Matches 50; Conservative 29; Mismatches 82; Indels 46; Gaps 8;  
 QY 1 MKTLKTLAVSASSLLASANAISYGNASDAQPIYGANIGVDANKQINGKNTAYGIAG 60  
 DB 1 MKNIIISTVITTSYLVVNAQADTNAFSGYARYAQSKV--ODFKNIGVNVKY----- 53  
 QY 61 YNDQNFGEAEFVGSDAKERNAGVSPKGDVKSFGA-----YGT----- 100  
 DB 54 -----RYEDSPVSVFSSISLYLGDQASGSVEPEGIHYHDKFEYKYSLWGP 102  
 QY 101 -YRYNFINTPFYAKGLGIATK-TKVDTSNATYTSNK--SDKTSIAGVGVGFKPLANV 156  
 DB 103 AYRLS-DNFSLYLAAGVGYATKFKHSHSTQDGDSPSKTKSSKRTGTGAWGAGVOMPLENI 161  
 QY 157 GVEASY---NYLSEDANAISLGAHLA 180  
 DB 162 VVDVGEYSGNSISSTKINGENVGVGYRF 188  
 RESULT 10  
 NMPC\_ECOLI STANDARD; PRT; 365 AA.  
 AC P21420; P77189;  
 DT 01-MAY-1991 (Rel. 18, Created)  
 DT 01-MAY-1991 (Rel. 18, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE OUTER MEMBRANE PORIN PROTEIN NMPC PRECURSOR.  
 GN NMPC OR PHMA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-MUTANT CS384;  
 RX MEDLINE-86304457; PubMed-3017988;

RA Blasband A.J., Marcotte W.R. Jr., Schnaltman C.A.;  
 RT "Structure of the Jc and mmpc outer membrane porin protein genes of  
 RL lambdaoid bacteriophage.";  
 RN J. Biol. Chem. 261:12723-12732(1986).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-K12 / MG1655;  
 RX MEDLINE-97426617; PubMed-9278503;  
 RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
 RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,  
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
 RA Mau B., Shao Y.;  
 RT "The complete genome sequence of Escherichia coli K-12.";  
 RL Science 277:1453-1474(1997).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RA Chung E., Allen E., Araujo R., Aparicio A., Davis K., Duncan M.,  
 RA Federapfel N., Hyman R., Kalman S., Komp C., Kordi O., Lew H., Lin D.,  
 RA Naneish A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
 RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 [4]  
 RP SEQUENCE OF 1-17 FROM N.A.  
 RC STRAIN-JL5502;  
 RX MEDLINE-94335635; PubMed-8057841;  
 RA Coll J.L., Heyde M., Portetier R.;  
 RT "Expression of the mmpc gene of Escherichia coli K-12 is modulated by  
 RT external pH. Identification of cis-acting regulatory sequences  
 RL Mol. Microbiol. 12:83-93(1994).  
 [5]  
 RP SEQUENCE OF 347-365 FROM N.A.  
 RC STRAIN-K12;  
 RX MEDLINE-96196428; PubMed-8648624;  
 RA Mandi A.A., Sharples G.J., Mandal T.N., Lloyd R.G.;  
 RT "Holiday junction resolvases encoded by homologous rusa genes in  
 RT Escherichia coli K-12 and phage 82.";  
 RL J. Mol. Biol. 257:561-573(1996).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.  
 CC -1- MISCELLANEOUS: IN WILD-TYPE STRAINS OF E. COLI K12, THE NMPC OPEN  
 CC READING FRAME IS INTERRUPTED BY AN IS5 INSERTION AND GENERATES A  
 CC HYBRID OPEN READING FRAME THAT IS NOT EXPRESSED. HOWEVER, IN  
 CC MUTANT STRAIN CS348, THE IS5 ELEMENT HAS BEEN DELETED AND NMPC IS  
 CC EXPRESSED.  
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHO FAMILY OF PORINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M13457; AAA3728.1; ALT\_SEQ.  
 DR EMBL: AE000160; AAC73654.1; ALT\_SEQ.  
 DR EMBL: U82598; AAB40749.1; ALT\_SEQ.  
 DR EMBL: Z35442; CA84594.1; -.  
 DR EMBL: X92587; CA63325.1; -.  
 DR PIR: A25647; MMECNC.  
 DR HSSP: P02931; 1GPN.  
 DR ECGENE: EG10659; NMPC.  
 DR INTERPRO: IPR001702; -.  
 DR INTERPRO: IPR001897; -.  
 DR PFM: PF00267; Gram-ve--porins; 1.  
 DR PRINTS: PR00182; ECOLNEIPORIN.  
 DR PRINTS: PR00183; ECOLIPORIN.  
 DR PROSITE: PS00576; GRAM\_NEG\_PORIN; 1.  
 KW Outer membrane; transmembrane; Porin; Signal.  
 FT SIGNAL 1 23  
 FT CHAIN 24 365 OUTER MEMBRANE PORIN PROTEIN NMPC.  
 FT CONFLICT 326 326 K -> N (IN REF. 2 AND 3).  
 FT SEQUENCE 365 AA; 40316 MW; 6E5128D4847FB4F8 CRC64;

Query Match 11.8%; Score 108; DB 1; Length 365;  
 Best Local Similarity 26.6%; Pred. No. 0.036;  
 Matches 62; Conservative 19; Mismatches 78; Indels 74; Gaps 11;  
 QY 1 MKTLKTL-AYSASSLLMSANAISY-----GNSDAQPYVCAK 39  
 DB 1 MKTLVAISAVAAVSVMASOAAEIYKNSKLDLYGVNAKHYFSSNDADGDTYAR 60  
 QY 40 IGVAVAKINGKNTFYAGYNFQNTGVFAEFGVSPAKFEFNAGVSPKDVKSFGAYG 99  
 DB 61 LGFKETQINDQITFGQWE-YEFKGN---RAESGSSKDKTRLAFAGLK-----FGDYG 111  
 QY 100 TYRYNFINTPEYAKGL-----GIKTRVDY-----TSRN----- 129  
 DB 112 SIDYGRNNGVAVDIGAMTDVLPFEFGDTWTQDVFMTORATGVATRYNNDFGLVDLNF 171  
 QY 130 APTYSNKSQDKT---SLAGVGVEFPLNANVVEASYNLSLSDANAISGAHLA 179  
 DB 172 AAQYQGNKDRSDFDNYTEGNGGF-----GFSATYEX-----EGFGIGATYA 213  
 RESULT 11  
 OM31\_BRUME STANDARD; PRT; 240 AA.  
 AC Q45322;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE 31 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.  
 GN OMP31.  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Brucellaceae; Brucella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-16M;  
 RX MEDLINE-96355886; PubMed-8751924;  
 RA Vicalino N., Cloeckaert A., Zygmunt M.S., Dubray G.;  
 RT "Cloning, nucleotide sequence, and expression of the Brucella  
 RT melitensis omp31 gene coding for an immunogenic major outer membrane  
 RT protein.";  
 RL Infect. Immun. 64:3744-3751(1996).  
 CC -1- FUNCTION: MAJOR OUTER MEMBRANE PROTEIN ASSOCIATED WITH  
 CC PEPTIDOGLYCAN. MAY FUNCTION AS A PORIN.  
 CC -1- SUBUNIT: OLIGOMERIC.  
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF076290; AAB36693.1; -.  
 DR INTERPRO: IPR000498; -.  
 DR PFM: PF01389; Ompa\_membrane; 1.  
 KW Antigen; Outer membrane; Porin; Signal.  
 FT SIGNAL 1 19  
 FT CHAIN 20 240 POTENTIAL.  
 FT DOMAIN 48 83 31 KDA OUTER-MEMBRANE IMMUNOGENIC  
 FT PROTEIN.  
 FT EPTROPE RECOGNIZED BY THE MONOCLONAL  
 FT ANTI-BODY A59/10F09/510.  
 FT SEQUENCE 240 AA; 25323 MW; 21C65EC479F66A5 CRC64;

Query Match 11.5%; Score 105; DB 1; Length 240;  
 Best Local Similarity 24.4%; Pred. No. 0.041;  
 Matches 51; Conservative 29; Mismatches 85; Indels 44; Gaps 9;

```
OY 4 KTLAVASASSLLAMSAANA---ISYNSADAP-----YVGAKIG----- 41
Db 1 MSVILASIAFATSAAMAADVVSSEAPADFTSWTGYIGIYAGGKFKHPF 60
OY 42 ----GVDAKOLNGK-----NTAVGIYAGYNDQNGV-----EAEFVSGDAK-ERNAGVS 86
Db 61 SSFDEDEDDVSGSLDVTAGCGFVGQVAGYNNQDNGVLAETDFQSSVTGSIASAGS 120
OY 87 PVKGDVKS-FGAYGTYRYNFINTP-----FYAKGLGIAGTKVDVT-SKNATTYSNKSDDK 139
Db 121 GLEGAEKFEKVEFGTVRRARLGTATERLMTVGTGGLANGKYSANLNGDDASALHTMSDK 180
OY 140 TSLAGVGVGFKPLANVGVEASYNVLSDD 168
Db 181 TKAGWTLGAGAEYAINNMWTLKSELYTD 209
```

```
RESULT 12
OM2B_HAEIN STANDARD: PRT; 363 AA.
AC Q48220:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-3230B;
RX MEDLINE=94018553; PubMed=8412618;
RA Duijn B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
non-encapsulated Haemophilus influenzae.";
RL Microb. Pathog. 14:451-462(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73391; CA51808.1; -
KM Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 363 OUTER MEMBRANE PROTEIN P2.
FT VARIANTS 21 363
SQ SEQUENCE 363 AA; 39873 MW; 2739FD0BF42A3051 CRC64;
```

```
Query Match 11.5%; Score 105; DB 1; Length 363;
Best Local Similarity 26.9%; Pred. No. 0.063;
Matches 52; Conservative 29; Mismatches 78; Indels 34; Gaps 11;
```

```
OY 5 KTLAVASASSLLAMSAANAISYNSADAPYVGAKIGY-----DAKQING-----K 51
Db 3 KTLAALIVGAFAPASANAANAAYVYNNEG-TNVELGGRISLIAEOSNSTTIKKOQKHGALRNQ 61
OY 52 NTAVGIYAGYNDQNGV-----GVEAEFVSGDAKEFNAGVSPVKGVSFGAYGT-YRYNFI 106
Db 62 SSRFHFKATHNFGDGYAGYLETREIVLSAQ-----SGTESD--NFGHITTKAYVYL 111
OY 107 NTPPYAKGLGIAGTKVD-VTSRNATTYS--NKSDDTSLAGC-VGVGFKPLANVGVEASY 162
Db 112 GNKACGEVILGKRAKTIADITSADKEYGVLNNSKYIPTNGWTVGTYTFKIDGLVLAGNY 171
OY 163 NYLSEDAANISIG 175
Db 172 -LLAQRNKYGTG 183
```

```
RESULT 13
OM25_HAEIN STANDARD: PRT; 371 AA.
AC P46027;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE OUTER MEMBRANE PROTEIN P2 PRECURSOR (OMP P2).
GN OMP2.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
CC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE=94293786; PubMed=8022287;
RA Duijn B., van Alphen L., Eljk P., Jansen H.M., Dankert J.;
RT "Antigenic drift of non-encapsulated Haemophilus influenzae major
outer membrane protein P2 in patients with chronic bronchitis is
caused by point mutations.";
RL Mol. Microbiol. 11:1181-1189(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T1, AND T2;
RX MEDLINE=94018553; PubMed=8412618;
RA Duijn B., Dankert J., Jansen H.M., van Alphen L.;
RT "Genetic analysis of the diversity in outer membrane protein P2 of
non-encapsulated Haemophilus influenzae.";
RL Microb. Pathog. 14:451-462(1993).
CC -----
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
-1- SIMILARITY: TO PROTOBACTERIUM STRAIN SS9 OMPH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73386; CA51803.1; -
DR EMBL; X73383; CA51800.1; -
KM Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 20
FT CHAIN 21 371
FT VARIANTS 184 224 K -> Q (IN T2).
FT VARIANTS 224 224 A -> T (IN T2).
FT VARIANTS 273 273 S -> T (IN T2).
SQ SEQUENCE 371 AA; 40901 MW; D0F252D7A6E5CDA6 CRC64;
```

```
Query Match 11.3%; Score 103; DB 1; Length 371;
Best Local Similarity 24.8%; Pred. No. 0.095;
Matches 53; Conservative 29; Mismatches 74; Indels 58; Gaps 12;
```

```
OY 5 KTLAVASASSLLAMSAANAISYNSADAPYVGAKIGY-----DAKQING-----K 51
Db 3 KTLAALIVGAFAPASANAANAAYVYNNEG-SKVELGGRISLIAEOSNNTVDDOQKHGALRNQ 61
OY 52 NTAVGIYAGYNDQNGV-----GVEAEFVG--SDAKEFNAGVSPVKGVSFGAYGTYRYN 104
Db 62 GSRFHFKATHNFGDGYAGYLETREISIQDNADHFD-----DITF-----KYAYV 108
OY 105 FINTPYAKGLGIAGTKV-VVTSRNATTYS--NKSDDTSLAGC-VGVGFKPLANVGVA 160
Db 109 TLGNKAFGEVKILGKRAKTIADITSADKEYGVLNNSKYIRTNCGWTVGTYTFKID--GLVL 166
OY 161 SYNLSEDA-----NAISIGA 176
Db 167 GANYLLAQRDTPANGPKGGEVAAGSISNGVOVCA 200
```

```
RESULT 14
ID OM25_BRUAB STANDARD: PRT: 213 AA.
AC 044664;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 25 KDA OUTER-MEMBRANE IMMUNOGENIC PROTEIN PRECURSOR.
GN OMP25.
OS Brucella abortus.
OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
OC Brucellaceae: Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-544S;
RX MEDLINE-95204367; PubMed=7896724;
RA de Wergifosse P., Lintermans P., Limet J.N., Cloeckaert A.;
RT "Cloning and nucleotide sequence of the gene coding for the major 25-
RL kDa outer membrane protein of Brucella abortus.";
CC J. Bacteriol. 177:1911-1914(1995).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE OMP25 / ROPB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: X79284; CAA55872.1; -.
CC INTERPRO: IPR000498; -.
CC DR PFAM: PF01389; OmpA-membrane; 1.
CC KW Antigen; Outer membrane; Signal.
CC FT SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 213 25 KDA OUTER-MEMBRANE IMMUNOGENIC
CC FT PROTEIN.
CC SQ SEQUENCE 213 AA; 23052 MW; 23285151F1F794BC7 CRC64;
CC -----
Query Match 11.2%; Score 102.5; DB 1; Length 213;
Best Local Similarity 27.1%; Pred. No. 0.058;
Matches 60; Conservative 22; Mismatches 60; Indels 79; Gaps 11;
QY 1 MKTLTLAVSSSLIAMSANAISYGNSSADA--QPYVGAKI----- 40
DB 1 MRTLKSLIVSA-ALLPESATAF-----AADAIOQPPVPAPEVAPQYSNAGTYTGLYL 54
QY 41 --GOVDAR-----QINGKNYAGIYAGYNFDON--FGEVAEFVGSDAKEFNAGVSPVKG 90
DB 55 GYGNMKAKTSYVGSIKIPDWMKRAGAFGWNFOODIYVGEVDAGYSMAKKSMDGLEVKGG 114
QY 91 DVKSGAAGTTRYNNINPFYAKKGLGT-----ATTKVDY 125
DB 115 FEGSLGAVGYDLNIV-MPYLTAGIAGSQIKLNLGDLDESKFRVWGWTAGAGLEAKLTNDI 173
QY 126 TSR---NATTSNKSDDKSTLAG-----GTVGVGFK 151
DB 174 LGRVEIRYTYGKNK--YDLAGTYRNKLDIOTDIRVIGIKY 212
CC -----
RESULT 15
VAC3_HELPY
ID VAC3_HELPY STANDARD: PRT: 1310 AA.
AC 048253;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE VACUOLATING CYTOTOXIN PRECURSOR.
GN VACA.
OS Helicobacter pylori (Campylobacter pylori).
```

```
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TX30A;
RX MEDLINE-95355366; PubMed=7629077;
RA Atherton J.C., Cao P., Peek R.M. Jr., Tummuru M.K., Blaser M.J.,
RA Cover T.L.;
RT "Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori.
RT Association of specific vacA types with cytotoxin production and
RT peptic ulceration.";
RL J. Biol. Chem. 270:17771-17777(1995).
CC -1- FUNCTION: INDUCES VACUOLATION OF EUKARYOTIC CELLS. CAUSES
CC ULCERATION AND GASTRIC LESIONS.
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL: U29401; AAA86834.1; -.
CC KW Cytotoxin; Toxin; Signal.
CC FT SIGNAL 1 30 POTENTIAL.
CC CHAIN 31 ? VACUOLATING CYTOTOXIN.
CC FT PROPEP ? 1310 POTENTIAL.
CC SQ SEQUENCE 1310 AA; 141988 MW; 1BC21FE3D435F981 CRC64;
CC -----
```

```
Query Match 11.2%; Score 102.5; DB 1; Length 1310;
Best Local Similarity 22.7%; Pred. No. 0.4;
Matches 48; Conservative 36; Mismatches 74; Indels 53; Gaps 10;
QY 9 AVSASSLLIAMSANAISYGNSSADAQPYVGAKI-----GOVDARQINCKN 52
DB 1049 AIGGTS-LNNGSNMSTL-YGTAGVDAYLNGEVAIVGGFGSYGSSFSFNQANSINSGANN 1106
QY 53 TAVGIYAGYNFDON-FGEVAE-FVGSDAKEFNAGVSPYKGVKSP-----GAGYT 100
DB 1107 TNEGYTSRIEFANQHEFDEAGAGLSDOSSLNFKSALLQDLNOSTHYIAYSATTTRASTG- 1165
QY 101 YRYNEINTPFYAKGKLGIAKTKVDVTSRNATTSNKSDDKSTIAGGVGFEKPLANVGEA 160
DB 1166 YDFAEFRNALVLKPSVGVSYNHL-----GSTNFKSNSQVALLNSGSSQHLFNNANVEA 1220
QY 161 SYN-----ISEDA-----NAISL 174
DB 1221 RYYGDTSYFYMNAGVLOEFARFGSNNAVSL 1251
```

Search completed: March 6, 2001, 14:40:05  
Job time: 173 sec







Db 2 MKMAPSLIAIAMAAGATAHAADDIYFGAGAAHFNGLNKIGGGYAGTEDDAAANA - 60  
 QY 52 NNAVYGYAGYNFQNGVE--AEFVG---SDAKEF-NAGVS-----PVKGPVKSFGAY 98  
 Db 61 -----FVYINTENFTEFGYQYAGRGNTDGLRINQCATISGLARLPLGGDFSAFAG 114  
 QY 99 GYRYNFIPTPYAKGLGIATKVDTSRNATVYSNKSDDKTSLAGVGVGEFPLANVCV 158  
 Db 115 GAY-----NAHFDGLGTSPTKV-----SPLAGLGVTYVNDALDL 149  
 QY 159 EASYNTLSEDAN 170  
 Db 150 QARYRYMVDVAD 161

RESULT 5  
 Q54470  
 ID 054470 PRELIMINARY; PRT; 172 AA.  
 AC Q54470;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE OUTER MEMBRANE PROTEIN PRECURSOR.  
 GN OMP4.  
 OS Serratia marcescens.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Serratia.  
 OX NCBI\_TaxID=615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=N28B;  
 RX MEDLINE=96036211; PubMed=7582013;  
 RA Guasch J.F., Ferrer S., Enfedaque J., Viejo M.B., Regue M.;  
 RT "A 17 kDa outer-membrane protein (Omp4) from Serratia marcescens  
 RT confers partial resistance to bacteriocin 28b when expressed in  
 RT Escherichia coli.";  
 RL Microbiology 141:2535-2542(1995).  
 DR EMBL: Z37157; CA85513.1; -.  
 DR HSSP: P36546; 10J8.  
 DR INTERPRO: IPR000498; -.  
 DR INTERPRO: IPR000758; -.  
 DR PFAM: PF01389; Omp4\_membrane; 1.  
 DR PRINTS: PRO0316; ENTEROVIROMP.  
 DR PROSITE: PS00694; ENT\_VIR\_OMP\_1; 1.  
 DR PROSITE: PS00695; ENT\_VIR\_OMP\_2; UNKNOWN\_1.  
 KW Signal; Outer membrane.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT CHAIN 25 172 OUTER MEMBRANE PROTEIN.  
 SQ SEQUENCE 172 AA; 18434 MW; 9E8283AAAE1994DF CRC64;

Query Match 12.9%; Score 118; DB 2; Length 172;

Best Local Similarity 25.0%; Pred. No. 0.015; Matches 48; Conservative 26; Mismatches 86; Indels 32; Gaps 7;

QY 1 MKTLKTLTLLAVSASSLLAMSAALISYNSADAPYVGAIGQVDAKQINKNATVAYGIAG 60  
 Db 1 MKTIACLSAVAACVL-----AVSAGTAPAGQSTVSAGTACGDFGVANKAGFNLKTR 53  
 QY 61 YNFDON-FGVAEAFVGSDAKEFNAGVSPVKGDFSGAYGTYRYNFIINT--PFY----- 111  
 Db 54 YEFDNPNPLGVIGSF-----THLEKDSQDGFYKKAQYNSISAGPAVRINDMAS 101  
 QY 112 AKKGLGIATKVVVTSRNATVYSNKSDDKTSLAGVGVGEFPLANVGEASY---NTLSED 168  
 Db 102 IYGLVGLGCGKFTTNQNGTNRHDTAD-YGFTYAGAGLQGFNPIENVALDVGEQNRIRSYD 160  
 QY 169 ANALISGAHLAF 180  
 Db 161 VGTWNVGVGYRF 172

RESULT 6

085182  
 ID 085182 PRELIMINARY; PRT; 572 AA.  
 AC 085182;  
 DT 01-NOV-1998 (TREMBLrel. 08, Created)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE FLAGELLIN A.  
 GN FLAA.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D2290;  
 RA Meinersmann R.J., Hielt K.L.;  
 RT "Concerted evolution of duplicate fla genes in Campylobacter";  
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: AF050188; AAC25641.1; -.  
 DR INTERPRO: IPR001029; -.  
 DR INTERPRO: IPR001492; -.  
 DR PFAM: PF00669; Flagellin\_N; 1.  
 DR PFAM: PF00700; Flagellin\_C; 1.  
 DR PRINTS: PR00207; FLAGELLIN.  
 DR PRODOM: PD000316; -; 1.  
 SQ SEQUENCE 572 AA; 59021 MW; C07AC87B421B59B1 CRC64;

Query Match 12.8%; Score 117; DB 2; Length 572;

Best Local Similarity 24.5%; Pred. No. 0.081; Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYNSMDAQYVCAKIGQVDAKO-----INCKNTAYGIYA----- 59  
 Db 242 AAVRAQATSDTFAINGVKTKGVKDGKGDANGALVAAINSVKDTTGVAEASIDANGQLITS 301  
 QY 60 ----GYNFDONFGVEAFVGSDAKEFNAGVSPVKGDFK-----SFGAYGTYRY 103  
 Db 302 REGRGIKIDONIGGGA-FITADMKENVGRSLVKNKGDKLLIGSNLSAGFGA----- 354  
 QY 104 NFINTPPYAKGLGIATK--VDVTSRNATVYSNKSDDKTSLAG-----GVGVGFKP 152  
 Db 355 ----TFISQASVSLRESKQOIDANIADAMGFGSANKGVVLGYSVAYSMSAGSGFSS 410  
 QY 153 LANVGEASYNVISEDANAISLGA 176  
 Db 411 GSGYSVSGSGKNYSTGFANALTAISA 434

RESULT 7

Q30689  
 ID 030689 PRELIMINARY; PRT; 572 AA.  
 AC 030689;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
 DE FLAGELLIN A.  
 GN FLAA.  
 OS Campylobacter jejuni.  
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
 OC Campylobacter.  
 OX NCBI\_TaxID=197;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=D1118;  
 RA Meinersmann R.J., Hielt K.L.;  
 RT "Concerted evolution of duplicate fla genes in Campylobacter";  
 RL Submitted (Feb-1998) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 101-189 FROM N.A.  
 RC STRAIN=D224;  
 RA Meinersmann R.J., Hiesel L.O., Fields P.I., Hielt K.L.;  
 J. Clin. Microbiol. 0:0-0(1997).



RESULT 10  
Q9R952 PRELIMINARY; PRT; 572 AA.  
AC Q9R952;  
DT 01-MAY-2000 (TREMblrel. 13, Created)  
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)  
DE FLAGELLIN A.  
GN FLAA.  
OS Campylobacter jejuni.  
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;  
OC Campylobacter  
OX NCBI\_TaxID=197;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=117;  
RA Meinersmann R.J., Hiett K.L.;  
RT "Concerted evolution of duplicate fla genes in Campylobacter.";  
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF050192; AAC25645.1; -  
DR INTERPRO: IPR001029; -  
DR INTERPRO: IPR001492; -  
DR PFAM: PF00669; Flagellin\_N; 1.  
DR PFAM: PF00700; Flagellin\_C; 1.  
DR PRINTS: PR00207; FLAGELLIN.  
SQ SEQUENCE 572 AA; 59038 MW; F59AA4104EBAZESE CRC64;

Query Match 12.6%; Score 115; DB 2; Length 572;  
Best Local Similarity 24.5%; Pred. No. 0.12;  
Matches 50; Conservative 24; Mismatches 70; Indels 60; Gaps 7;

QY 22 AATSYGNSADAQPYVGAKIGOVDAKQ-----INGKNATYGIYA----- 59  
DB 242 AAVRAGATSDPRAINGVKIGVDYKGDANGALVAAINSVKDTTGEASIDANGOLLIS 301  
QY 60 ---GYNDONFGEAEFVGSDAKEFNAGVSPYKGVK-----SEFAGYGYRY 103  
DB 302 RECRGKIKDNTGCGA-FINADMKENYGRSLVKNDGKILISGSLMSAGFEA----- 354  
QY 104 NFNTPFYAKKIGLAKTR--VDVTSRNATTTYSNKSDDKTSLAG-----GVGCVFPR 152  
DB 355 ---TFISQASVSLRESGQIDANIDAMGFGSANKGVGLGYSSVAYMSAGSGFSS 410  
QY 153 LANVGEASVNTLSEDAATISLGA 176  
DB 411 GSGYVSGSKNYSSTGFANAIATISA 434  
RESULT 11  
Q9KS75 PRELIMINARY; PRT; 160 AA.  
AC Q9KS75;  
DT 01-OCT-2000 (TREMblrel. 15, Created)  
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)  
DE HYPOTHETICAL PROTEIN VC1384.  
GN VC1384.  
OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=EL TOR N16961 / SEROTYPE O1;  
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,  
Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,  
Gill S.R., Nelson K.B., Read T.D., Tettelin H., Richardson D.,  
Ermondlaev M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,  
McDonald L., Uitterback T., Fleischmann R.D., Nieman W.C., White O.,  
Salberg S.L., Smith H.O., Colwell R.R., Mekalanos J.D., Venter J.C.,  
Fraser C.M.;  
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.";  
RL Nature 406:477-483(2000).  
DR EMBL; AE004217; AAF94542.1; -  
DR TIGR; VC1384; -  
KW Hypothetical protein.  
SQ SEQUENCE 160 AA; 16985 MW; 7E47CED047CE409D CRC64;

Query Match 12.6%; Score 114.5; DB 2; Length 160;  
Best Local Similarity 29.8%; Pred. No. 0.026;  
Matches 56; Conservative 19; Mismatches 70; Indels 43; Gaps 9;

QY 5 KTLAVASASSLLAMSANAAISYGNSSADAQPYVGAKIGOVDAKQINGKNATYGIYAGYND 64  
DB 3 KTLAL--ALLGASSTA-----MADSMYGGASVGGSDPECKHG--TAYSVHACTGIL 51  
QY 65 QNFGVAEEFVGSDAKEFNA-----GVSPYKDVKSFPACIGYRINFTNTPYAKGL---- 116  
DB 52 PFILGELAGYVNHGDFEINATQELASSLYFAVFKPSMDFGLH-----VYAKGLHSMD 104  
QY 117 -GIATKRVDTSRNATTTYSNKSDDKTSLAGVGVGFKPLANVGEASY-NYL--SEDAAT 172  
DB 105 KDINGSKID-----DGDVMTYIGAEVFTIGFPVSGASYMTYMDSTDVCTL 151  
QY 173 SLGAHLAF 180  
DB 152 SFNATFHF 159

RESULT 12  
Q46679 PRELIMINARY; PRT; 250 AA.  
AC Q46679;  
DT 01-NOV-1996 (TREMblrel. 01, Created)  
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)  
DE HEAT RESISTANT AGGLUTININ 1.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OX Escherichia  
RN NCBI\_TaxID=562;  
RP SEQUENCE FROM N.A.  
RC STRAIN=O9:H10:K99;  
RX MEDLINE=95012721; PubMed=7927783;  
RA Lutyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;  
RT "Cloning, sequencing, and viscometric adhesion analysis of heat-  
resistant agglutinin 1, an integral membrane hemagglutinin from  
Escherichia coli O9:H10:K99.";  
RT Infect. Immun. 62:5020-5026(1994).  
DR EMBL: U07174; AAC13752.1; -  
DR INTERPRO: IPR000498; -  
DR PFAM: PF01389; OmpA\_membrane; 1.  
FT CHAIN 26  
SQ SEQUENCE 250 AA; 27477 MW; BDDAE89C389BD395 CRC64;

Query Match 12.0%; Score 109.5; DB 2; Length 250;  
Best Local Similarity 21.0%; Pred. No. 0.12;  
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

QY 1 MKTKTLAVASASSLLAMSANAAIS-----VGNSSADAQ 34  
DB 1 MEMKNVIAVSALAMGMSSTALADESKTGFVYTKACASVNSLADQRLSCNGEETSK 60  
QY 35 YVGAKIGOVDAKQINGKNATY--GIYAGVNFQDNFGV---EAEFV--GSDAKEFNAGVS 86  
DB 61 YKGG-----DGHDTVSGGIAAGVDEYQFSIPVTELEFAARGKADSKYVWDK 110  
QY 87 PYVG-----DVKS-----FGAYGYRINFTNTPYAKKIGLAK-----TKVDV-- 125  
DB 111 SMSGYWRDCLKNEVSVNTLMLNAYYDFRNSAFTWVSAG-IGYAKELHQTGTISTWD 169

OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNLSEDAANAIS 173  
DB 170 YGCGSSGRESLSRSGSADNFAMSLAGVRYDVTPTDIALDLSTRYLDAGDSSVS 222

RESULT 13  
ID 046678 PRELIMINARY: PRT: 264 AA.  
AC 046678:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE HEAT RESISTANT AGGLUTININ 1 PRECURSOR.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O9:H10-K99;  
RX MEDLINE-95012721; PubMed-7927783;  
RA Lulwyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;  
"Cloning, sequencing, and viscometric adhesion analysis of heat-  
resistant agglutinin 1, an integral membrane hemagglutinin from  
RT Escherichia coli O9:H10-K99."  
RL Infect. Immun. 62:5020-5026(1994).  
DR EMBL: U07174; AAC13751.1; -.  
DR INTERPRO: IPR000498; -.  
DR PFAM: PF01389; OmpA\_membrane; 1.  
KW Signal.  
FT SIGNAL 1 86 POTENTIAL.  
FT CHAIN 40 264 HEAT RESISTANT AGGLUTININ 1.  
SQ SEQUENCE 264 AA: 29024 MW; 335BDAE8E2D4F67 CRC64;

Query Match 12.0%; Score 109.5; DB 2; Length 264;  
Best Local Similarity 21.0%; Pred. No. 0.13;  
Matches 49; Conservative 38; Mismatches 75; Indels 71; Gaps 10;

OY 1 MTKLTLLAVSASSLLANSANAIS-----YGNSDAQP 34  
DB 15 MIEMKVIIVSALAMAGMSTQALADESKTGFYTGKAGASVMSIADQRLSGNGEETSK 74  
OY 35 YGCATIGQVDAQKQINKTAY--GIYAGNFDQNGV-----EAERV--GSDAKKEFNAGVS 86  
DB 75 YKGG-----DGHDTVFSGGIAAGDYFPOESIPVTELEFYARGKADSKYNVDKD 124  
OY 87 PVKG-----DVKS-----FGAYGTYRYNFINTPFYAKGKLGIAK-----TKVDV-- 125  
DB 125 SMSGGYWRDLDLNEVSNTLMNLATYDFRNDSAFTFWVSAG--IGYAKELHQKTTGISTWD 183  
OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNLSEDAANAIS 173  
DB 184 YGCGSSGRESLSRSGSADNFAMSLAGVRYDVTPTDIALDLSTRYLDAGDSSVS 236

RESULT 14  
ID 033796 PRELIMINARY: PRT: 165 AA.  
AC 033796:  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
DE ATTACHMENT AND INVASION PROTEIN HOMOLOG.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmonella.  
OX NCBI\_TaxID=602;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA De Groote M., Ochsen U.A., Shiloh M., Nathan C., McCoard J.M.,  
Dinauer M.C., Libby S.J., Vazquez-Torres A., Xu Y., Fang F.C.;  
Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF007380; AAB62386.1; -.  
DR HSSP: P36546; 1008.  
SQ SEQUENCE 165 AA: 17439 MW; 4F998FE466A15B88 CRC64;

Query Match 12.0%; Score 109; DB 2; Length 165;  
Best Local Similarity 25.6%; Pred. No. 0.076;  
Matches 41; Conservative 19; Mismatches 68; Indels 32; Gaps 6;

OY 36 VGAKIGQVDAKQING-KMT-----AYGIYAGYNE---DQNFGEAEFV 74  
DB 10 VLALGSLGVRNAAGYKNTVSTGIAVYTDLSGLSGNANCAKIKYMWEDLDSGFGAMGSYT 69  
OY 75 GSDAKKEFNAGVSPVKDVKFSFGAYGTYRYN-FINTPFYAKKLGIAKRYDVTSHNATY 133  
DB 70 YTSADVNNGYKVGADADYTSLLVGSYRFNDYLN-----AYVIGAANGHI-----KDNW 119

OY 134 SNKSDKTSIAGVGGEFKPLANVGVEASYNLSEDAANAIS 173  
DB 120 GNSDNKTAFAYGAGIGIOLNPEVNIIVNASYEHTSFSTDADS 159

RESULT 15  
ID 046680 PRELIMINARY: PRT: 247 AA.  
AC 046680:  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
DE HEAT RESISTANT AGGLUTININ 1.  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-O9:H10-K99;  
RX MEDLINE-95012721; PubMed-7927783;  
RA Lulwyche P., Rupp R., Cavanagh J., Warren R.A., Brooks D.E.;  
"Cloning, sequencing, and viscometric adhesion analysis of heat-  
resistant agglutinin 1, an integral membrane hemagglutinin from  
RT Escherichia coli O9:H10-K99."  
RL Infect. Immun. 62:5020-5026(1994).  
DR EMBL: U07174; AAC13753.1; -.  
DR INTERPRO: IPR000498; -.  
DR PFAM: PF01389; OmpA\_membrane; 1.  
FT CHAIN 23 247 HEAT RESISTANT AGGLUTININ 1.  
SQ SEQUENCE 247 AA: 27104 MW; FFE5D1B261802389 CRC64;

Query Match 11.9%; Score 108.5; DB 2; Length 247;  
Best Local Similarity 20.9%; Pred. No. 0.14;  
Matches 48; Conservative 38; Mismatches 73; Indels 71; Gaps 10;

OY 4 LKTLTLLAVSASSLLANSANAIS-----YGNSDAQP 37  
DB 1 MNKVIIVSALAMAGMSTQALADESKTGFYTGKAGASVMSIADQRLSGNGEETSKYG 60  
OY 38 AKIGQVDAKQINKTAY--GIYAGNFDQNGV-----EAERV--GSDAKKEFNAGVSPVK 89  
DB 61 G-----DGHDTVFSGGIAAGDYFPOESIPVTELEFYARGKADSKYNVDKDSWS 110  
OY 90 G-----DVKS-----FGAYGTYRYNFINTPFYAKKLGIAK-----TKVDV----- 125  
DB 111 GGYWRDLDLNEVSNTLMNLNAYDFRNDSAFTFWVSAG--IGYAKELHQKTTGISTWDYGY 169  
OY 126 -----TSRNATYTSNKSDDKTSIAGVGGEFKPLANVGVEASYNLSEDAANAIS 173  
DB 170 GSSGRESLSRSGSADNFAMSLAGVRYDVTPTDIALDLSTRYLDAGDSSVS 219

Search completed: March 6, 2001, 14:38:28

Job time: 131 sec

---

**THIS PAGE BLANK (USPTO)**